

1974

Estimation of regression parameters for finite populations

Michael Arsene Hidirolou
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Hidirolou, Michael Arsene, "Estimation of regression parameters for finite populations " (1974). *Retrospective Theses and Dissertations*. 5146.
<https://lib.dr.iastate.edu/rtd/5146>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

- 1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.**
- 2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.**
- 3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.**
- 4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.**
- 5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.**

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

75-10,481

HIDIROGLOU, Michael Arsene, 1948-
ESTIMATION OF REGRESSION PARAMETERS
FOR FINITE POPULATIONS.

Iowa State University, Ph.D., 1974
Statistics

Xerox University Microfilms, Ann Arbor, Michigan 48106

Estimation of regression parameters
for finite populations

by

Michael Arsene Hidioglou

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1974

TABLE OF CONTENTS

	Page
I. INTRODUCTION AND REVIEW OF LITERATURE	1
II. SOME LARGE SAMPLE THEORY	14
A. Order in Probability	14
B. Convergence in Distribution	22
C. Some Central Limit Theorems	25
D. Approximation to the Expectation	26
III. FOUR APPROACHES TO THE LIMITING DISTRIBUTIONS IN SIMPLE RANDOM SAMPLING FROM A FINITE POPULATION: APPLICATIONS TO REGRESSION	30
A. Madow's Condition W	30
B. Erdős-Rényi's Condition	35
C. Hajek's Condition	37
D. A Central Limit Theorem for Regression in a Finite Population Using Hajek's Theorem	59
E. A Superpopulation Approach to the Limiting Distribution of Regression Coefficients in a Finite Population	78
IV. ERRORS-IN-VARIABLES	88
A. Introduction	88
1. The model	88
B. An Errors-in-variables Regression Model for Clusters	92
C. Model for Clustered Data Subject to Error	96
1. Introduction	96
2. The model and assumptions	98
3. Estimation	102
4. Variance of the regression coefficients given normal error structure	108
5. An estimator for the variance of the regression coefficients given non- normal error structure	122

	Page
D. Estimation when the Error Variance is a Multiple of Total Variance	133
1. Introduction	133
2. The model and assumptions	134
3. Estimation	135
4. Variance of the regression coefficients given normal error structure	141
5. An estimator for the variance of the regression coefficients given non- normal error structure	149
6. A modified estimator	157
E. An Example	161
V. A MONTE CARLO STUDY	186
VI. SUMMARY	204
VII. LITERATURE CITED	212
VIII. ACKNOWLEDGMENTS	218

I. INTRODUCTION AND REVIEW OF LITERATURE

In recent years a considerable number of books on the design and analysis of sample surveys have been published. Examples are Deming [1950], Hansen, Hurwitz and Madow [1953], Sukhatme [1954], Yates [1949], Cochran [1963] and Kish [1965]. Characteristics of the theory presented in these books are as follows:

- 1) the sample population contains a finite number of elements,
- 2) no assumptions are made concerning the distributions of the pertinent variables in the population,
- 3) major emphasis is placed on the estimation of simple population parameters such as percentages, means and totals,
- 4) the sample sizes are assumed to be "large". This allows the sampling distributions of estimates to be approximated by normal distributions.

Sample surveys where major emphasis is placed on the estimation of population parameters, such as percentages, means or totals, have been called descriptive or enumerative surveys. An analytical survey is one where comparisons are made between different subgroups of the population. The interest of an analytic survey is the formulation or testing of hypothesis about forces at work in the population.

Simple random sampling without replacement is the base for the existing body of sample survey theory. For a number of reasons, modifications of this sampling method are often necessary. For instance, one rarely samples from a finite population without having

some prior knowledge about the population. This prior information can often be used in the sample design to increase the precision of estimators. The available techniques are stratification, post-stratification, selection with probabilities proportional to some auxiliary variable, ratio estimation and regression estimation. Since many finite populations are composed of natural clusters or groups of elements, cost considerations lead to the use of multi-stage sampling procedures.

In cases where the selection of elements is without replacement or (and) in clusters, a dependence among observations is introduced. As a result, estimators such as regression estimators require approximate procedures for evaluating their variances.

A major question in the analysis of analytical surveys concerns the conceptual view of the finite population. The finite population can be considered to be a fixed set of elements, or to be a sample from an infinite population. In discussing the latter view, Fisher [1928, p. 700] stated: "The idea of an infinite hypothetical population is, I believe, implicit in all statements involving mathematical probability."

Cochran [1939] was one of the first survey statisticians to consider the superpopulation concept. He viewed the finite population as being sampled from an infinite population with finite first and second moments, and used this concept to compare the relative precision of various sampling schemes.

In recent years, a number of studies have appeared in the survey

literature which explicitly considers the finite population to be a sample from an infinite population. J. N. K. Rao [1973] provides a bibliography of these studies. When the population displays definite "non-normal" characteristics, infinite population models have been used as a justification for estimators other than the mean, [Brewer and Ferrier, 1966; Fuller, 1970]. Also they have been used as a justification for certain sampling or estimation procedures [Royall, 1970; Kalbfleisch and Sprott, 1969] and as a basis for survey design [Ericson, 1969; Isaki, 1970].

Consider a superpopulation where the random variable of interest is Y . We denote the expected value and variance of Y by E_S and V_S respectively. The subscript "S" stands for superpopulation operations.

Assume that

$$E_S(Y) = \mu$$

$$V_S(Y) = E_S(Y - \mu)^2 = \sigma^2 .$$

First, a random sample of size N is drawn from this superpopulation. Denote the selected elements by Y_1, Y_2, \dots, Y_N . Second, a simple random sample of size n is drawn without replacement (W.O.R.) from the chosen population. Denote the selected sample elements by y_1, y_2, \dots, y_n .

Let the mean and variance of the finite population be \bar{Y} and S^2 respectively, where

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 ,$$

and let the mean and variance of the sample be \bar{y} and s^2 respectively, where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i ,$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 .$$

Using the symbols " E_F " and " V_F " to denote the expected value and variance of the sample estimate over all possible samples of size n selected from the finite population,

$$E_F(\bar{y}) = \bar{Y}$$

$$V_F(\bar{y} - \bar{Y}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n} .$$

Denoting the expectation over all finite populations of N elements by E and the variance by V ,

$$E\{\bar{y}\} = E_S(E_F\bar{y})$$

$$= E_S(\bar{Y})$$

$$= \mu ,$$

$$V\{\bar{y}\} = E_S(V_F(\bar{y})) + V_S(E_F(\bar{y}))$$

$$= E_S\left(\frac{N-n}{Nn} S^2\right) + V_S(\bar{Y})$$

$$= \frac{N-n}{Nn} E_S(S^2) + V_S(\bar{Y})$$

$$= \frac{\sigma^2}{n} .$$

When using \bar{y} to estimate the superpopulation mean μ , there is no finite population correction (f.p.c.) in the variance, but when using y to estimate the finite population mean \bar{Y} , there is a finite population correction in the variance. This point has been made by Deming [1950, p. 251] and Cochran [1963, p. 37]. Cochran says, in reference to the comparison of two subpopulation means: "One point should be noted. It is seldom of scientific interest to ask whether $\bar{Y}_j = \bar{Y}_k$ because these means would not be exactly equal in a finite population, except by a rare chance, even if the data in both domains were drawn at random from the same infinite population. Instead, we test the null hypothesis that the two domains were drawn from infinite populations having the same mean. Consequently we omit

the f.p.c. when computing $V(\bar{y}_j)$ and $V(\bar{y}_k)$. . ."

We have introduced the conceptual problems faced by the sampler when conducting an analytical survey. When the surveys are complex there are also technical problems raised by the analytical use of data. These problems often involve ordinary regression techniques. Considering these problems, Cochran [1963] states:

The theory of linear regression plays a prominent part in statistical methodology. The standard results of this theory are not entirely suitable for sample surveys because they require the assumptions that the population regression of y on x is linear, that the residual variance of y about the regression line is constant, and that the population is infinite. If the first two assumptions are violently wrong, a linear regression estimate will probably not be used. However, in surveys in which the regression of y on x is thought to be approximately linear, it is helpful to be able to use \bar{y}_{lr} without having to assume exact linearity or constant residual variance.

Consequently we present an approach that does not demand that the regression in the population be linear. The results hold only in large samples. They are analogous to the large-sample theory for the ratio estimate.

Hartley [1959] arrives at a similar conclusion in his paper on analyses for domains of study. He says: " . . . Nevertheless we shall not employ regression estimators. The reason for this is not that we consider regression theory inappropriate, but that the theory for finite populations requires considerable development before it can be applied in the present situation."

Sedransk argued that analytical work with survey data should be done "by design", [Sedransk, 1965]. That is, areas and methods of analysis should be set forth before taking the survey, and the sample should be selected to conform as closely as possible to the requirements of the stated methods. Sedransk [1965] assumed that

the primary goal of an analytical survey is to compare the means of different domains of study. If \hat{Y}_i and \hat{Y}_j are the estimated means for the i^{th} and j^{th} domains, Sedransk places constraints on the variance of their difference for all i and j , and searches for sample-size allocations that will minimize cost functions. A variety of different situations are considered: 1) random samples selected from each of the domains, 2) a random sample selected from the overall population, 3) two-stage cluster samples selected from each of the domains, and 4) two-stage cluster samples selected from the total population. In the second and fourth cases, Sedransk considers double sampling procedures. To satisfy constraints phrased in terms of all possible pairwise domain comparisons, approximate solutions for the sample sizes are obtained.

Of interest is Sedransk's [1965] comment: "Often the inference desired from the sample is to relate to a more 'general' population than that represented by a finite one from which the sample was physically selected. Moreover, other (unknown) persons will use the sample data to make inferences for other (finite or infinite) populations. In such situations the model to be used must represent the true population of interest as accurately as possible." Fuller [1973] had similar views: "In the regression studies of sample data falling within our personal experience, the investigator was interested in conclusions beyond the finite population actually sampled. This does not mean that the investigator could specify the population of finite populations that have been or will be generated by the "mechanism"

that generated the finite population under current study." When the investigator wishes to broadest possible inferences, he should choose a model with the potential for generalization. Fuller [1973] says:

Treating the finite population as a sample from an infinite population is one framework which provides the potential for generalization. In fact, we believe a strong case can be made for the following position: "The objective of a regression study of survey data is the construction and estimation of a linear model such that the sample data are consistent with the hypothesis that the data are a random sample from an infinite population wherein the linear model holds . . ."

Deming and Stephan [1941] consider a sample drawn from a given population as being sampled from a superpopulation, the given population in question being one of an infinity of possible populations. Hence, assuming that our data arises as a result of some underlying system subject to chance variables, it is acceptable to use the present sample for predictive purposes. For example, Fuller and Battese [1973] presented a linear regression model that was used in the analysis of data from a longitudinal regression model for Ames women. In the analysis, they concluded that the height of these women decreases as age increases. It may be of interest to know how well this model holds for women at different times and places. The following comment by Leon Truesdell (referred to by Deming and Stephan, 1941) may be helpful: "A so-called 100 percent sample from the viewpoint of scientific method is as soon taken, a sample of the past. The usefulness of such a sample is only as a basis for drawing an inference about the future and in this case the sample is but a finite sample of a potentially infinite one that might result from the cause system existing at the time the sample was taken."

Konijn [1962] introduced a regression model in which he considered estimators of the coefficients of a regression equation for the population. It was assumed that the surveyed population consisted of M individuals belonging to N classes. Konijn assumed that "the M individuals arose as a proportionate stratified sample from an infinitely large population of individuals with similar behavior."

It was assumed that in each of the N strata of the conceptual population we have the regression model

$$y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij} \quad i=1, 2, \dots, N; j=1, 2, \dots, M_i$$

$$E(y_{ij} | x_{ij} = x_o) = \alpha_i + \beta_i x_o$$

$$E(e_{ij} | x_{ij} = x_o) = 0$$

$$E(e_{ij}^2 | x_{ij} = x_o) = \sigma_i^2.$$

Konijn defined a weighted average for α and β ,

$$\alpha = \left[\sum_{i=1}^N M_i \right]^{-1} \left[\sum_{i=1}^N M_i \alpha_i \right]$$

$$\beta = \left[\sum_{i=1}^N M_i \right]^{-1} \left[\sum_{i=1}^N \beta_i \right].$$

Konijn felt that if the β_i differed very much from each other as compared to σ_i^2 , there was little point in estimating β . A simple random sample of size n is drawn without replacement from the N

clusters. From each selected cluster i of size M_i , a subsample of size m_i is drawn using random sampling without replacement.

Konijn suggested the unbiased estimators of α and β ,

$$\hat{\alpha} = \left[\frac{1}{N} \sum_{i=1}^N M_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n M_i \hat{\alpha}_i \right]$$

$$\hat{\beta} = \left[\frac{1}{N} \sum_{i=1}^N M_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n M_i \hat{\beta}_i \right]$$

where

$$\hat{\alpha}_i = \bar{y}_{i.} - \hat{\beta}_i \bar{x}_{i.}$$

$$\hat{\beta}_i = \frac{\sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{i.}) y_{ij}}{\sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{i.})^2}$$

$$\bar{x}_{i.} = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij}$$

$$\bar{y}_{i.} = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij} .$$

The variance of $\hat{\beta}$ is given by

$$V(\hat{\beta}) = \frac{n}{N} \left[\sum_{i=1}^N \sigma_{i.k_i}^2 + \frac{N-n}{N-1} \sum_{i=1}^N \left(\beta_{i.k_i} - \frac{\sum_{i=1}^N \beta_{i.k_i}}{N} \right)^2 \right] ,$$

where

$$F_i^2 = E_i \left[\frac{1}{m_i \sum_{j=1} (x_{ij} - \bar{x}_{i.})^2} \right]$$

E_i = conditional variance given strata i

$$k_i = \left[\frac{n}{N} M \right]^{-1} M_i .$$

Konijn constructed an unbiased estimator of $V(\hat{\beta})$ given by

$$\hat{V}(\hat{\beta}) = \frac{n}{N} \left[\sum_{i=1}^n \frac{\frac{\hat{\sigma}_i^2 k_i^2}{m_i \sum_{j=1} (x_{ij} - \bar{x}_{i.})^2}}{\sum_{j=1} (x_{ij} - \bar{x}_{i.})^2} + \frac{N-n}{N-1} \sum_{i=1}^n (\hat{\beta}_i k_i - \frac{\hat{\beta}}{n})^2 \right]$$

where

$$\hat{\sigma}_i^2 = \frac{1}{m_i - 2} \sum_{j=1}^{m_i} [(y_{ij} - \bar{y}_{i.})^2 - \hat{\beta}_i (x_{ij} - \bar{x}_{i.}) y_{ij}] .$$

The variance of $\hat{\beta}$ may be interpreted as the sum a term associated with the variability of $\hat{\beta}_i$ as an estimator of β_i and a term due to the sampling variability of the chosen β_i 's.

Frankel [1971] conducted an empirical investigation on data collected by the U.S. Bureau of the Census in the March 1967 Current Population Survey. This monthly survey has a stratified cluster-sample design. Frankel studied the empirical behavior of estimators

for means, correlation coefficients, multiple correlation coefficients and partial correlation coefficients.

Frankel considered the estimation of regression-type parameters defined by

$$B = (X_N' X_N)^{-1} X_N' Y_N$$

where N denotes the number of units in the surveyed population; X_N denotes the $(N \times p)$ matrix of observations on p -independent variables for all units of the population; and Y_N denotes the $(N \times 1)$ vector of observations on the dependent variable of interest.

Frankel investigated the properties of the estimator

$$b = (X_n' X_n)^{-1} X_n' Y_n ,$$

where n denotes the number of units selected in the sample; X_n denotes the $(n \times p)$ matrix of observations on the p independent variables for the sampled units; and Y_n denotes the $(n \times 1)$ vector of observations on the dependent variable of interest. The estimates of variance for his estimators were computed using Taylor expansions in terms of population moments, balanced repeated replication (B.R.R.) and jack-knife repeated replication (J.R.R.).

From his Monte-Carlo study, Frankel concluded that for each parameter of interest;

- A. The sample estimator is approximately unbiased for the population parameter.

- B. An approximately unbiased estimator of the variance of the sample estimator is computable from the sample.
- C. The distribution of the ratio of the sample estimator minus its expected value to its estimated standard error is approximated by the Student-t-distribution.

It should be noted that Frankel defined the regression estimates in terms of the finite population values.

II. SOME LARGE SAMPLE THEORY

In this chapter, we shall review the concepts of relative magnitude or order. These concepts were introduced by Mann and Wald [1943] and generalized by Chernoff [1956]. These concepts are useful in investigating the limiting behavior of random variables.

The definitions and theorems are presented as in Fuller [1972].

A. Order in Probability

Let $\{a_n\}$ be a sequence of real numbers and $\{r_n\}$ a sequence of positive real numbers.

Definition 2.A.1

We say a_n is of order $o(r_n)$,

$$a_n = o(r_n)$$

if

$$\lim_{n \rightarrow \infty} \frac{a_n}{r_n} = 0 .$$

Definition 2.A.2

We say a_n is of order $O(r_n)$

$$a_n = O(r_n)$$

if for some finite real number M

$$\frac{|a_n|}{r_n} \leq M$$

for all n .

Definition 2.A.3

The sequence of random variables $\{X_n\}$ converges in probability to the random variable X and we write

$$\text{plim } X_n = X$$

or

$$X_n \xrightarrow{P} X ,$$

if for every $\epsilon > 0$ and $\delta > 0$, there exists an N such that for $n > N$

$$P \{ |X_n - X| > \epsilon \} < \delta .$$

We now define order in probability.

Definition 2.A.4

Let $\{X_n\}$ be a sequence of random variables and $\{r_n\}$ a sequence of positive real numbers. We say that X_n is of probability $O_p(r_n)$ and write

$$X_n = O_p(r_n)$$

if for every $\epsilon > 0$ there exists a positive real number M_ϵ and an N_ϵ such that

$$P \{ |X_n| \geq M_\epsilon r_n \} \leq \epsilon$$

for all $n > N_\epsilon$.

We say that X_n is of probability order $o_p(r_n)$ and write $X_n = o_p(r_n)$ if

$$\text{plim} \left(\frac{X_n}{r_n} \right) = 0.$$

We now proceed to state a theorem by Mann and Wald [1943] which shows that the algebra (definitions 2.A.1 and 2.A.2) extends to the algebra of the order in probability relationships (definition 2.A.4).

Theorem 2.A.1

Let $\{X_n\}$ be a sequence of k -dimensional random variables with element $x_n^{(j)}$, $j=1, 2, \dots, k$ and $g_n(X_n)$ be a sequence of measurable functions. Let $\{s_n\}$ and $\{r_{jn}\}$ be $k+1$ sequences of positive numbers. If

$$x_n^{(j)} = o_p(r_n^{(j)}) \quad j=1, 2, \dots, t$$

$$x_n^{(j)} = o_p(r_n^{(j)}) \quad j = t+1, t+2, \dots, k$$

and if for any nonrandom sequence $\{a_n\}$ such that

$$g_n(a_n) = o(s_n)$$

whenever

$$a_n^{(j)} = o(r_n^{(j)}) \quad j=1, 2, \dots, t$$

$$a_n^{(j)} = o(r_n^{(j)}) \quad j=t+1, \dots, k$$

then

$$g_n(x_n) = O_p(a_n) .$$

Proof:

(See Fuller [1972].)

Corollary 2.A.1

If $\{X_n\}$ is a sequence of scalar random variables and

$$x_n = a + o_p(r_n)$$

where $r_n \rightarrow 0$ and if $g(x)$ is a function with a continuous derivative at $x=a$, then

$$g(x_n) = g(a) + \sum_{j=1}^{s-1} \frac{1}{j!} g^{(j)}(a) (x_n - a)^j + o_p(r_n^s) ,$$

where $g^{(s)}(a)$ is the s^{th} derivative of $g(x)$ evaluated at $x=a$.

If $o_p(r_n)$ is replaced by $O_p(r_n)$, $o_p(r_n^s)$ is correspondingly replaced by $O_p(r_n^s)$.

Proof:

(See Fuller [1972].)

Using the definitions and properties of limits, it is easily shown that

i) If $a_n = o(f_n)$ and $b_n = o(g_n)$

then

$$a_n b_n = o(f_n g_n)$$

$$(a_n + b_n) = o\{\text{Max}(f_n, g_n)\}$$

ii) If $a_n = O(f_n)$ and $b_n = O(g_n)$

then

$$a_n b_n = O(f_n g_n) .$$

Now, if we have sequences of random variables $\{X_n\}$ and $\{Y_n\}$ such that

$$X_n = o_p(f_n) \quad \text{and} \quad Y_n = o_p(g_n) ,$$

it follows by Theorem 2.A.1 that

$$X_n Y_n = o_p(f_n g_n)$$

$$(X_n + Y_n) = o_p\{\text{Max}(f_n, g_n)\} .$$

We now state the Tchebychef inequality.

Theorem 2.A.2

If X is a random variable with finite variance and distribution function $F(x)$, then for every $\epsilon > 0$ and finite A

$$P(|X-A| \geq \epsilon) \leq \frac{E(X-A)^2}{\epsilon^2} .$$

Proof:

(See Fuller [1972].)

Corollary 2.A.2

Given the sequence of random variables $\{X_n\}$ satisfying

$$E\{X_n^2\} = O(a_n^2)$$

then

$$X_n = O_p(a_n) .$$

Proof:

(See Fuller [1972].)

Corollary 2.A.3

Given $s > 0$ and the sequence of random variables $\{X_n\}$ satisfying

$$E\{(X_n - E(X_n))^2\} = O(a_n^2)$$

$$E\{X_n\} = O(a_n)$$

then

$$X_n^S = O_p(a_n^S) .$$

Proof:

(See Fuller [1972].)

Lemma 2.A.1

If X and Y are k -dimensional random variables then for every

$\epsilon > 0$

$$P\{|X=Y| \geq \epsilon\} \leq P\{|X| \geq \frac{\epsilon}{2}\} + P\{|Y| \geq \frac{\epsilon}{2}\} .$$

Proof:

(See Fuller [1972].)

Theorem 2.A.3

If the sequence X_n and Y_n is such that

$$\lim_{n \rightarrow \infty} E\{(X_n - Y_n)^2\} = 0$$

$$\text{plim } X_n = Y$$

then

$$\text{plim } Y_n = Y .$$

Proof:

(See Fuller [1972].)

Corollary 2.A.4

If the sequence of random variables X_n is such that

$$\lim_{n \rightarrow \infty} E(X_n) = \mu$$

$$\lim_{n \rightarrow \infty} E\{(X_n - E(X_n))^2\} = 0$$

then

$$\text{plim } X_n = \mu .$$

Proof:

(See Fuller [1972].)

Theorem 2.A.4

If $g: R^{k_1} \longrightarrow R^{k_2}$ (R^{k_1} and R^{k_2} denoting Euclidean k_1 and k_2 dimensional spaces respectively) is a continuous function and $\underline{X}_n \xrightarrow{P} \underline{X}$ where \underline{X}_n is a k_1 dimensional vector, then

$$g(\underline{X}_n) \xrightarrow{P} g(\underline{X})$$

where $g(\underline{X}_n)$ is k_2 dimensional.

Proof:

Let $\epsilon_1 > 0$ and $\epsilon_2 > 0$ be given. Let A be a finite closed k_1 dimensional rectangle such that $P(\underline{X} \in A) \geq 1 - \frac{\epsilon_2}{2}$. g continuous on its domain implies g is uniformly continuous on A . Let $d[\underline{X}_n, \underline{X}]$

denote the distance between the points \underline{X}_n and \underline{X} in R^k , and let $d[g(\underline{X}_n), g(\underline{X})]$ denote the distance between the points $g(\underline{X}_n)$ and $g(\underline{X})$ in R^{k_2} .

Then, there exists $\delta > 0$ such that $\underline{X} \in A$ and $d[\underline{X}_n, \underline{X}] < \delta$ and $d[g(\underline{X}_n), g(\underline{X})] < \epsilon_1$. Therefore

$$\begin{aligned} P(d[g(\underline{X}_n), g(\underline{X})] \geq \epsilon_1) &\leq P(\underline{X} \notin A) + P(d[\underline{X}_n, \underline{X}] \geq \delta) \\ &\leq \frac{\epsilon_2}{2} + P(d[\underline{X}_n, \underline{X}] \geq \delta) . \end{aligned}$$

Since $\text{plim } \underline{X}_n = \underline{X}$, there exists an N such that, for $n \geq N$

$$P(d[\underline{X}_n, \underline{X}] \geq \delta) \leq \frac{\epsilon_2}{2} .$$

The theorem therefore follows. //

B. Convergence in Distribution

Definition 2.B.1

Given $\{X_n\}$ a sequence of random variables with distribution functions $\{F_n\}$. X_n is said to converge in distribution to the random variable X with distribution function F , denoted by

$$X_n \xrightarrow{L} X$$

if $\lim_{n \rightarrow \infty} F_n = F$

at every continuity point of F .

Theorem 2.B.1

If $\text{plim } |X_n - Y_n| = 0$

$$X_n \xrightarrow{L} X$$

then

$$Y_n \xrightarrow{L} X .$$

Proof:

(See Fuller [1972].)

Corollary 2.B.2

If $\text{plim } X_n = X$

then

$$X_n \xrightarrow{L} X .$$

That is, convergence in probability implies convergence in distribution.

Proof:

(See Fuller [1972].)

Corollary 2.B.3

If $g(X)$ is a continuous function except on a set D where

$$P(X \in D) = 0$$

and

$$\text{plim } \bar{X}_n = X$$

then

$$g(X_n) \xrightarrow{L} g(X).$$

Proof:

(See Fuller [1972].)

Theorem 2.B.2

Let $\{X_n\}$ be a sequence of random variables such that F_n is the distribution function of X_n and let X be a random variable with distribution function F .

If $F_n \longrightarrow F$ then

$$\int g dF_n \longrightarrow \int g dF$$

for every bounded continuous function g .

Proof:

(See Rao [1965], page 97.)

Theorem 2.B.3

Let $\{X_{1n}\}$, $\{X_{2n}\}$, ... be sequences of random variables converging in probability to the constraints c_1 , c_2 , ... respectively and let $g(X_{1n}, X_{2n}, \dots)$ be any rational function. Then

$$g(X_{1n}, X_{2n}, \dots) \xrightarrow{P} g(c_1, c_2, \dots)$$

if $g(c_1, c_2, \dots) < \infty$.

Proof:

(See Cramér [1946], page 255.)

C. Some Central Limit Theorems

Theorem 2.C.1 (Multivariate Central Limit Theorem)

Let F_n denote the joint distribution function of the k dimensional random variable $(X_{1n}, X_{2n}, \dots, X_{kn})$, $n=1, 2, \dots$. Let F be the joint distribution function of the k dimensional random variable (X_1, X_2, \dots, X_k) . Let $F_{\lambda n}$ be the d.f. of the linear function

$$\sum_{i=1}^k \lambda_i X_{in}.$$

Then the necessary and sufficient condition that

$$F_n \longrightarrow F$$

is that $F_{\lambda n} \longrightarrow F_{\lambda}$, for each arbitrary vector $(\lambda_1, \lambda_2, \dots, \lambda_k)$.

Proof:

(See Rao [1965], page 108.)

Theorem 2.C.2 (Liapounov Central Limit Theorem)

Let $\{X_n\}$ be a sequence of independent random variables, with finite $(2+\delta)^{\text{th}}$ moments, for some $\delta > 0$. If

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E\{(X_i - \mu_i)^{2+\delta}\} = 0$$

where

$$s_n^2 = \sum_{i=1}^n \sigma_i^2$$

$$\sigma_i^2 = E(X_i - \mu_i)^2$$

$$\mu_i = E(X_i)$$

then

$$\sum_{i=1}^n \left(\frac{X_i - \mu_i}{s_n} \right) \xrightarrow{L} N(0,1) .$$

Proof:

(See Breiman [1968].)

D. Approximation to the Expectation

A random variable X_n , may converge in probability and hence in distribution to a random variable X , the latter possessing finite moments even though $E\{X_n\}$ is not defined. On the other hand, it may be known that X_n has finite moments of order r , and

be interested in approximating the sequence defined by $E\{X_n\}$.

DeGracie and Fuller [1972] give the necessary conditions to specify that the expectation of the random variable differs from a specified sequence by an amount of specified order.

Theorem 2.D.1 (DeGracie-Fuller)

Let $\{X_n\}$ be a sequence of k dimensional random variables with distribution function $F_n(x)$ and let $\{f_n(x)\}$ be a sequence of functions mapping R^k into R . If

$$i) \int ||x-\mu||^{2s} dF_n(x) = O(a_n^{2s}) \text{ where } a_n = o(1)$$

$$ii) \int |f_n(x)|^2 dF_n(x) = O(1)$$

$$iii) \frac{\partial^s}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_s}} f_n(x) \text{ is continuous over a closed}$$

and bounded sphere S for all $n > N_0$

iv) μ is an interior point of S

v) There is a K such that for every $n > N$.

$$\left| \frac{\partial^2}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_s}} f_n(x) \right| \leq K \quad \text{for all } x \in S$$

$$\left| \frac{\partial^r}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_r}} f_n(\mu) \right| \leq K \quad \text{for } r=1, 2, \dots, s-1$$

$$|f_n(\mu)| \leq K$$

then

$$\begin{aligned} \int f_n(x) dF_n(x) &= f_n(\mu) + \sum_{j=1}^{s-1} \frac{1}{j!} \int D^j f_n(\mu) (x-\mu)^j dF_n(x) \\ &\quad + O(a_n^s) \end{aligned}$$

where

$$\begin{aligned} D^r f_n(\mu) (x-\mu)^r &= \sum_{i_1=1}^k \sum_{i_2=1}^k \dots \sum_{i_r=1}^k \frac{\partial^r}{\partial x_{i_1} \dots \partial x_{i_r}} f_n(\mu) \\ &\quad \cdot \prod_{j=1}^r (x_{i_j} - \mu_{i_j}) \end{aligned}$$

and for $s=1$ it is understood that

$$\int f_n(x) dF_n(x) = f_n(\mu) + O(a_n)$$

Proof:

(See Fuller [1972].)

Theorem 2.D.2 (DeGracie-Fuller)

Let $f_n(x)$ be a sequence of real valued functions and $\{X_n\}$ a sequence of k -dimensional random variables with distribution functions $F_n(x)$. If

i) $|f_n(x)| \leq K$ for $x \in \bar{S}$ where S is a bounded open set containing μ ,

ii) $|f_n(x)|^2 \leq Y(x)n^p$ for some $p > 0$ where

$$|Y(x)|^2 dF_n(x) = o(1)$$

iii) $\int ||x-\mu||^{4p} dF_n(x) = o(n^{-2p})$

then

$$\int |f_n(x)|^2 dF_n(x) = o(1) .$$

Proof:

(See Fuller [1972].)

III. FOUR APPROACHES TO THE LIMITING DISTRIBUTIONS IN SIMPLE RANDOM SAMPLING FROM A FINITE POPULATION: APPLICATIONS TO REGRESSION

In this chapter, we will discuss four approaches for obtaining the limiting distribution of the sample mean, based on a simple without replacement sample selected from a finite population. To study the asymptotic properties of the sample mean, a sequence of populations must be considered. The approaches considered are: Madow's [1948] Condition W, Erdős-Rényi's [1959] Condition, Hájek's [1960] Condition and the superpopulation model. We generalize the study of limiting distribution of sample means to the study of limiting distribution of regression coefficients.

A. Madow's Condition W

Madow [1948] considered the limiting distribution of the sample mean for a sample selected without replacement from a finite universe. He proved that the limiting distribution of $n^{\frac{1}{2}}(\bar{y}_n - \bar{Y}_N)$, where \bar{y}_n is the sample mean and \bar{Y}_N is the population mean is normal provided:

- i) as the size of the universe increases, the higher moments do not increase too rapidly relative to the variance, and
- ii) for sufficiently large sample and population sizes, the ratio of sample size to population size is bounded away from 1.

Madow defined a sequence of universes $\phi_1, \phi_2, \dots, \phi_r, \dots, \phi_r$ containing N_r elements, $\{Y_{ri}\}$ $i=1, 2, \dots, N_r$. A simple random sample of size n_r is drawn without replacement from ϕ_r : the selected elements are denoted by $\{y_{ri}\}$ $i=1, 2, \dots, n_r$.

The linear function

$$z_{n_r} = n_r^{-1/2} \left[\sum_{i=1}^{n_r} (y_{ri} - \bar{Y}_{N_r}) \right] \quad (3.A.1)$$

where

$$\bar{Y}_{N_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} Y_{ri}$$

is of interest.

The sampling variance of z_{n_r} , $\sigma_{z_{n_r}}^2$, is

$$\sigma_{z_{n_r}}^2 = \frac{N_r}{N_r - 1} \left(1 - \frac{n_r}{N_r} \right) \mu_{2N_r}, \quad (3.A.2)$$

where

$$\mu_{kN_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^k \quad (3.A.3)$$

$k=1, 2, \dots$. We will also use the symbol $\sigma_{Y_r}^2$ for μ_{2N_r} .

Madow assumed the following for the sequence of universes $\phi_1, \phi_2, \dots, \phi_r, \dots$.

Condition W

For sufficiently large n_r and N_r ,

$$i) \frac{n_r}{N_r} < 1-\epsilon, \text{ where } \epsilon > 0$$

and

ii) there exists a finite value λ , such that for all k ,

$$|\lambda_k(N_r)| < \lambda,$$

where

$$\lambda_k(N_r) = \frac{\mu_{kN_r}}{\left[\mu_{2N_r}\right]^{\frac{k}{2}}}.$$

Condition W is sufficient to prove the following theorem.

Theorem 3.A.1 (Madow)

If the sequence ϕ_r satisfies Condition W and a simple random sample of size n_r is selected without replacement from ϕ_r , then for all t

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r \rightarrow \infty}} P\{Z_{n_r} < t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{w^2}{2}} dw$$

where

$$z_{n_r} = \frac{z_{n_r}}{(\sigma_{z_{n_r}}^2)^{\frac{1}{2}}} \quad (3.A.4)$$

and z_{n_r} was defined in (3.A.1). Madow extended Theorem (3.A.1) to samples selected from multivariate populations.

Theorem 3.A.2 (Madow)

Suppose that the elements of ϕ_r are p-component vectors $Y_{ri.} = (Y_{ri1}, Y_{ri2}, \dots, Y_{rip}), i=1, 2, \dots, N_r$.

Assume that Condition W is satisfied for each component of this vector.

As before, let

$$z_{n_r j} = n_r^{-\frac{1}{2}} \sum_{i=1}^{n_r} (y_{rij} - \bar{Y}_{N_r j}) \quad j=1, 2, \dots, p \quad (3.A.5)$$

where y_{rij} is the i^{th} sample element and

$$\bar{Y}_{N_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} Y_{rij} . \quad (3.A.6)$$

Define

$$z_{n_r j} = \frac{z_{n_r j}}{(\sigma_{z_{n_r j}}^2)^{\frac{1}{2}}} , \quad (3.A.7)$$

where

$$\sigma_{z_{n_r j}}^2 = \frac{N_r}{N_r - 1} \left(1 - \frac{n_r}{N_r}\right) \mu_{2N_r j}$$

and

$$\mu_{kN_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{rij} - \bar{Y}_{N_r j})^k \quad k=1, 2, \dots$$

Suppose that $\lim_{r \rightarrow \infty} \rho_{rlj} = \rho_{lj}$ is defined for all l and j , where

$$\rho_{rlj} = \frac{\frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{ril} - \bar{Y}_{N_r l})(Y_{rij} - \bar{Y}_{N_r j})}{(\mu_{2N_r l})^{\frac{1}{2}} (\mu_{2N_r j})^{\frac{1}{2}}}$$

$$l, j = 1, 2, \dots, p \quad \text{for } l \neq j$$

and

$$\rho_{lj} > -1 + \epsilon, \epsilon > 0.$$

Then the limiting distribution of $(Z_{n_r 1}, Z_{n_r 2}, \dots, Z_{n_r p})$ is multivariate normal, with mean zero and covariance matrix Φ where

$$\Phi = \begin{bmatrix} 1, & \rho_{12}, & \dots, & \rho_{1p} \\ \rho_{21}, & 1, & \dots, & \rho_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{p1}, & \rho_{p2}, & \dots, & 1 \end{bmatrix}.$$

B. Erdős-Rényi's Condition

Erdős-Rényi [1959] investigated the limiting distribution of a sum of weakly dependent random variables, given that these variables were created as a without replacement sample selected from a finite population. As did Madow, he considered a sequence of finite populations $\phi_1, \phi_2, \dots, \phi_r, \dots$ such that the r^{th} population contained N_r elements.

Erdős-Rényi considered the sum

$$T_{n_r} = \sum_{i=1}^{n_r} y_{ri} \quad (3.B.1)$$

for all $\binom{N_r}{n_r}$ possible samples. Denote by $C_{N_r, n_r}(t)$ the number of sums T_{n_r} which do not exceed

$$\frac{n_r}{N_r} \sum_{i=1}^{N_r} Y_{ri} + t[V(T_{n_r})]^{1/2} \quad (3.B.2)$$

where

$$V(T_{n_r}) = \frac{n_r}{N_r} \frac{N_r - n_r}{N_r - 1} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2$$

and t is an arbitrary positive real number. Then

$$F_{n_r, n_r}(t) = \frac{C_{N_r, n_r}(t)}{\binom{N_r}{n_r}} \quad (3.B.3)$$

is the fraction of samples exceeding the specified limit.

Conditions on the sequence $\{Y_{ri}\}$, $i=1, 2, \dots, N_r$ in order that $F_{N_r, n_r}(t)$ converges to the normal $N(0,1)$ distribution function are stated in the following theorem.

Theorem 3.B.1 (Erdős-Rényi)

Consider a sequence of finite populations $\phi_1, \phi_2, \dots, \phi_r, \dots$ and let

$$d_{N_r, n_r}(\epsilon) = \frac{1}{V(T_{N_r})} \sum_{\substack{|y'_{ri}|^2 \\ |y'_{ri}| > \epsilon[V(S_{n_r})]^{1/2}}} \quad (3.B.4)$$

$$1 \leq i \leq N_r$$

where

$$T_{N_r} = \sum_{i=1}^{N_r} Y_{ri}$$

$$V(T_{N_r}) = \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2$$

$$y'_{ri} = y_{ri} - \bar{Y}_{N_r}$$

and $\epsilon > 0$ is an arbitrary real number.

If $n_r \leq N_r/2$ is chosen in such a manner that when $n_r \rightarrow \infty$ and $N_r - n_r \rightarrow \infty$, we have for any $\epsilon > 0$

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r - n_r \rightarrow \infty}} d_{N_r, n_r}(\epsilon) = 0, \quad (3.B.5)$$

then it follows that for any real t ,

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r - n_r \rightarrow \infty}} F_{N_r, n_r}(t) \xrightarrow{L} \Phi(t)$$

where $\Phi(t)$ is the normal distribution with mean zero and variance one.

C. Hájek's Condition

When sampling from finite populations without replacement, the sampled elements are not independent random variables, and hence, the Lindeberg condition [Rao, 1965, p. 107] cannot be directly applied to the study of limiting distributions for sums of these sampled elements. Hájek's contribution [1960] was the introduction of Poisson sampling to obtain the limiting distribution of the sample mean of these sampled elements.

Poisson sampling can be described as an experiment consisting of N_r dichotomous experiments. The population element i is either included in the sample or it is not included with probabilities

$$\frac{n_r}{N_r} \quad \text{and} \quad 1 - \frac{n_r}{N_r}$$

respectively. This is analogous to flipping a biased coin with the probability of a head equal to

$$\frac{n_r}{N_r} .$$

If the outcome is a head we include element i in the sample, if the outcome is a tail we exclude the element. The size of the Poisson sample which we denote by k_r is a binomial (Bernoulli) random variable attaining the value b with probability

$$\binom{N_r}{k_r} \left(\frac{n_r}{N_r} \right)^{k_r} \left(1 - \frac{n_r}{N_r} \right)^{N_r - k_r}, \quad 0 \leq k_r \leq N_r .$$

The mean value of the number of elements included is:

$$\begin{aligned} E(k_r) &= \sum_{i=0}^{N_r} i \binom{N_r}{i} \left(\frac{n_r}{N_r} \right)^i \left(1 - \frac{n_r}{N_r} \right)^{N_r - i} \\ &= n_r , \end{aligned} \tag{3.C.1}$$

and the variance of the number of elements is

$$\begin{aligned}
 E(k_r - n_r)^2 &= \sum_{i=0}^{N_r} (i - n_r)^2 \binom{N_r}{i} \left(\frac{n_r}{N_r} \right)^i \left(1 - \frac{n_r}{N_r} \right)^{N_r - i} \\
 &= n_r \left(1 - \frac{n_r}{N_r} \right) .
 \end{aligned} \tag{3.C.2}$$

As before let the sequence of populations be denoted by $\phi_1, \phi_2, \dots, \phi_r, \dots$. From ϕ_r , a Poisson sample of mean size n_r is selected. We assume that $n_r < n_{r+1}$ and $N_r < N_{r+1}$. Furthermore,

$$\frac{n_r}{N_r} < 1 - \epsilon \quad \text{where } \epsilon > 0 .$$

Let $s_{k_r} = \{y_{r1}, y_{r2}, \dots, y_{rk_r}\}$ be the r^{th} sample obtained using Poisson sampling.

Define

$$C_r = \sum_{i=1}^{n_r} (y_{ri} - \bar{Y}_{N_r}) \tag{3.C.3}$$

and

$$C_r^* = \sum_{i \in s_{k_r}} (y_{ri} - \bar{Y}_{N_r}) \tag{3.C.4}$$

where

$$\bar{Y}_{N_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} Y_{ri} .$$

Now C_r^* can be expressed as the sum of N_r independent variables c_{ri} , $i=1, 2, \dots, N_r$ where

$$P[c_{ri} = y_{ri} - \bar{Y}_{N_r}] = \frac{n_r}{N_r}$$

$$P[c_{ri} = 0] = 1 - \frac{n_r}{N_r} .$$

We are interested in the limiting distribution of C_r . The following Lemma enables us to establish that the limiting distributions of C_r and C_r^* are the same.

Lemma 3.C.1 (Hájek)

The following inequality holds:

$$\frac{E(C_r - C_r^*)^2}{V(C_r^*)} \leq \left(\frac{1}{n_r} + \frac{1}{N_r - n_r} \right)^{\frac{1}{2}} . \quad (3.C.5)$$

Proof:

Denote by s_{n_r} the set of elements drawn from ϕ_r using a simple random sampling scheme w.o.r., that is

$$s_{n_r} = \{y_{r1}, y_{r2}, \dots, y_{rn_r}\} .$$

Then $s_{n_r} \cap s_{k_r}^c$ and $s_{n_r}^c \cap s_{k_r}$ represent a simple random sample w.o.r. of size $|k_r - n_r|$.

$$\begin{aligned}
 V(C_r - C_r^*) &= E[V(C_r - C_r^*) | k_r] + V[E(C_r - C_r^*) | k_r] \\
 &= E[V(C_r - C_r^*) | k_r] \\
 &= E[E(C_r - C_r^*)^2 | k_r] .
 \end{aligned} \tag{3.C.6}$$

$$\begin{aligned}
 E[(C_r - C_r^*)^2 | k_r] &= \frac{|k_r - n_r|}{N_r} \frac{N_r - |k_r - n_r|}{N_r - 1} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2 \\
 &\leq |k_r - n_r| \sigma_{Yr}^2 .
 \end{aligned} \tag{3.C.7}$$

Hence

$$\begin{aligned}
 E[E(C_r - C_r^*)^2 | k_r] &\leq E|k_r - n_r| \sigma_{Yr}^2 \\
 &\leq [E(k_r - n_r)^2]^{1/2} \sigma_{Yr}^2 \\
 &= [n_r (1 - \frac{n_r}{N_r})]^{1/2} \sigma_{Yr}^2 .
 \end{aligned} \tag{3.C.8}$$

$$\begin{aligned}
 V(C_r^*) &= E[V(C_r^*) | k_r] + V[E(C_r^*) | k_r] \\
 &= E[V(C_r^*) | k_r] .
 \end{aligned} \tag{3.C.9}$$

Since s_{k_r} is a simple random sample w.o.r. from ϕ_r ,

$$V(C_r^* | k_r) = \frac{k_r}{N_r} \frac{N_r - n_r}{N_r - 1} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2 . \quad (3.C.10)$$

Hence using (3.C.1) and (3.C.2)

$$V(C_r^*) = n_r \left(1 - \frac{n_r}{N_r}\right) \sigma_{Yr}^2 , \quad (3.C.11)$$

where

$$\sigma_{Yr}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2$$

and

$$\begin{aligned} \frac{E(C_r - C_r^*)^2}{V(C_r^*)} &\leq \left[n_r \left(1 - \frac{n_r}{N_r}\right)\right]^{-1/2} \\ &= \left(\frac{1}{n_r} + \frac{1}{N_r - n_r}\right)^{1/2} . \quad // \end{aligned}$$

Lemma 3.C.2

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r \rightarrow \infty}} \frac{V(C_r^*)}{V(C_r)} = 1 . \quad (3.C.12)$$

Proof:

$$\frac{V(C_r^*)}{V(C_r)} = \frac{n_r^2 \left(\frac{1}{n_r} - \frac{1}{N_r}\right) \frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2}{n_r^2 \left(\frac{1}{n_r} - \frac{1}{N_r}\right) \frac{1}{N_r-1} \sum_{i=1}^{N_r} (Y_{ri} - \bar{Y}_{N_r})^2} .$$

The result follows by letting $n_r \rightarrow \infty$ and $N_r \rightarrow \infty$. //

From Lemma 3.C.1, it follows that

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r - n_r \rightarrow \infty}} \frac{E(C_r - C_r^*)^2}{V(C_r^*)} = 0 . \quad (3.C.13)$$

Relations (3.C.12) and (3.C.13) imply that, provided $n_r \rightarrow \infty$ and $N_r - n_r \rightarrow \infty$, the limiting variances and distributions of the random variables

$$\frac{A_r + B_r C_r}{[V(C_r)]^{\frac{1}{2}}}$$

and

$$\frac{A_r + B_r C_r^*}{[V(C_r^*)]^{\frac{1}{2}}}$$

exist under the same conditions, and if they exist they are the same. The random variable C_r^* , however, is a sum of independent random variables, so that when studying the limiting distributions of

$$\frac{A_r + B_r C_r^*}{[V(C_r^*)]^{\frac{1}{2}}},$$

we may apply the well-known theory for the sum of independent random variables.

We next state without proof Hájek's Central Limit Theorem.

Theorem 3.C.4 (Hájek)

Given a sequence of finite populations ϕ_r , $r=1, 2, \dots$, let $\phi_{r\tau}$ be the subset of elements of ϕ_r on which the inequality

$$|Y_{ri} - \bar{Y}_{N_r}| > \tau [V(C_r)]^{\frac{1}{2}}$$

holds, where $\tau > 0$ is an arbitrary real positive number.

Letting

$$Z_{n_r} = \frac{C_r}{[V(C_r)]^{\frac{1}{2}}}$$

then for any real t

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r - n_r \rightarrow \infty}} P\{Z_{n_r} < t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{w^2}{2}} dw$$

if and only if

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r - n_r \rightarrow \infty}} \frac{\sum_{i \in \phi_{rT}} (Y_{ri} - \bar{Y}_{N_r})^2}{N_r} = 0 . \quad (3.C.14)$$

Proof:

The proof follows directly from the Lindeberg Central Limit Theorem. //

We now show that Madow's Condition W is a stronger condition than Hájek's condition.

Theorem 3.C.5

Given that

$$\mu_{kN_r} = \mu_{2N_r}^{\frac{k}{2}} \lambda_k(N_r), \quad k = 2+\delta ,$$

for some $\delta > 0$ and

$$|\lambda_k(N_r)| < \lambda ,$$

then

$$\lim_{\substack{n_r \rightarrow \infty \\ N_r - n_r \rightarrow \infty}} \frac{\sum_{i \in \phi_{rT}} (Y_{ri} - \bar{Y}_{N_r})^2}{N_r} = 0 .$$

Proof:

Since

$$\phi_{r\tau} = \{Y_{ri} : |Y_{ri} - \bar{Y}_{N_r}| > \tau [V(C_r)]^{\frac{1}{2}}\}$$

we have

$$0 < \frac{1}{|Y_{ri} - \bar{Y}_{N_r}|} < \frac{1}{\tau [V(C_r)]^{\frac{1}{2}}}, \quad (3.C.15)$$

for $Y_{ri} \in \phi_{r\tau}$.

Now

$$V(C_r) = f_r (1 - f_r) \frac{N_r^2}{N_r - 1} \sigma_{Yr}^2$$

where

$$f_r = \frac{n_r}{N_r}.$$

Hence, using (3.C.15) we obtain

$$\frac{\sum_{i \in \phi_{r\tau}} |Y_{ri} - \bar{Y}_{N_r}|^2}{N_r} = \frac{\sum_{i \in \phi_{r\tau}} |Y_{ri} - \bar{Y}_{N_r}|^2}{N_r \sigma_{Yr}^2}$$

$$\begin{aligned}
&= \frac{1}{N_r \sigma_{Yr}^2} \sum_{i \in \phi_{rT}} \frac{|Y_{ri} - \bar{Y}_{N_r}|^{2+\delta}}{|Y_{ri} - \bar{Y}_{N_r}|^\delta} \\
&\leq \frac{1}{\tau^\delta} \frac{1}{[f_r(1-f_r)]^{\frac{\delta}{2}}} \frac{1}{N_r^{1+\frac{\delta}{2}}} \sum_{i \in \phi_{rT}} \left| \frac{Y_{ri} - \bar{Y}_{N_r}}{\sigma_{Yr}} \right|^{2+\delta} \\
&\leq \frac{1}{\tau^\delta} \frac{1}{[f_r(1-f_r)]^{\frac{\delta}{2}}} \frac{1}{N_r^{1+\frac{\delta}{2}}} \sum_{i=1}^{N_r} \left| \frac{Y_{ri} - \bar{Y}_{N_r}}{\delta_{Yr}} \right|^{2+\delta} \\
&< \frac{1}{\tau^\delta} \frac{1}{[f_r(1-f_r)]^{\frac{\delta}{2}}} \frac{1}{N_r^{\frac{\delta}{2}}} \lambda .
\end{aligned}$$

The result follows by letting $N_r \rightarrow \infty$ and $n_r \rightarrow \infty$ with

$$f_r = \frac{n_r}{N_r}$$

bounded away from one. //

We now proceed to develop a multivariate extension to Theorem 3.C.4. Let $\phi_1, \phi_2, \dots, \phi_r, \dots$ be a sequence of finite populations. A simple random sample of size n_r is selected without replacement from ϕ_r . Assume that $n_r > n_{r-1}$, $N_r > N_{r-1}$ and

$\frac{n_r}{N_r} < 1 - \epsilon$ for all r , where $\epsilon > 0$. The p -dimensional vector

$$Y_{ri\cdot} = (Y_{ri1}, Y_{ri2}, \dots, Y_{rip})$$

is associated with the i^{th} element of the r^{th} population. Define the normalized vector

$$Z_{ri\cdot} = \left(\frac{Y_{ri1} - \bar{Y}_{N_r 1}}{\sigma_{Yr1}}, \frac{Y_{ri2} - \bar{Y}_{N_r 2}}{\sigma_{Yr2}}, \dots, \frac{Y_{rip} - \bar{Y}_{N_r p}}{\sigma_{Yrp}} \right)$$

where

$$\bar{Y}_{N_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} Y_{rij}, \quad j=1, 2, \dots, p$$

$$\sigma_{Yrj}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{rij} - \bar{Y}_{N_r j})^2.$$

We need the following Lemma for the multivariate extension.

Lemma 3.C.3

Assume that

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in \mathcal{B}_j} Z_{rij}^2 = 0 \quad \text{for } j=1, 2, \dots, p \quad (3.C.16)$$

where $\tau > 0$ is some arbitrary positive real number and

$$\mathcal{B}_j = \{i \in \mathcal{I}_r : z_{rij}^2 > \tau n_r\} .$$

Then condition (3.C.16) holds if and only if

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in \mathcal{A}} \|z_{ri.}\|^2 = 0 \quad (3.C.17)$$

where

$$\|z_{ri.}\|^2 = \sum_{j=1}^p z_{rij}^2$$

and

$$\mathcal{A} = \{i \in \mathcal{I}_r : \|z_{ri.}\|^2 > \tau n_r\} .$$

Proof:

We first prove that condition (3.C.16) implies condition (3.C.17). Suppose

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in \mathcal{B}_j} z_{rij}^2 = 0 \quad \text{for } j=1, 2, \dots, p .$$

Now,

$$\mathcal{A} \subset \mathcal{C}$$

where

$$C = \bigcup_{j=1}^p C_j$$

and

$$C_j = \{i \in \phi_r: z_{rij}^2 > \frac{\tau_{nr}}{p}\} .$$

Hence

$$\frac{1}{N_r} \sum_{i \in A} \|z_{ri.}\|^2 \leq \frac{1}{N_r} \sum_{i \in C} \|z_{ri.}\|^2$$

$$= \frac{1}{N_r} \sum_{i \in C} \sum_{j=1}^p z_{rij}^2$$

$$= \frac{1}{N_r} \sum_{\substack{p \\ i \in \bigcup_{j=1}^p C_j}} \sum_{j=1}^p z_{rij}^2$$

$$\leq \frac{1}{N_r} \sum_{l=1}^p \left[\sum_{i \in C_l} \sum_{j=1}^p z_{rij}^2 \right]$$

$$= \sum_{l=1}^p \left[\frac{1}{N_r} \sum_{i \in C_l} \sum_{j=1}^p z_{rij}^2 \right] .$$

(3.C.18)

Also

$$\frac{1}{N_r} \sum_{i \in C_1} \sum_{j=1}^P z_{rij}^2 = \frac{1}{N_r} \sum_{i \in C_1} z_{ril}^2 + \frac{1}{N_r} \sum_{i \in C_1} \sum_{\substack{m=1 \\ (m \neq 1)}}^P z_{rim}^2 \quad (3.C.19)$$

and

$$\frac{1}{N_r} \sum_{i \in C_1} z_{rim}^2 \leq \frac{1}{N_r} \sum_{i \in C_1} z_{ril}^2 + \frac{1}{N_r} \sum_{i \in C_m} z_{rim}^2. \quad (3.C.20)$$

Substituting (3.C.19) and (3.C.20) into (3.C.18) we obtain,

$$\frac{1}{N_r} \sum_{i \in A} \|z_{ri.}\|^2 \leq (2p-1) \sum_{l=1}^P \frac{1}{N_r} \sum_{i \in C_l} z_{ril}^2.$$

Letting $r \rightarrow \infty$, we obtain that

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in A} \|z_{ri.}\|^2 = 0.$$

We next prove that condition (3.C.17) implies condition (3.C.16). Suppose

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in A} \|z_{ri.}\| = 0 \quad (3.C.21)$$

Define

$$\mathcal{B} = \bigcup_{j=1}^p \mathcal{B}_j$$

where

$$\mathcal{B}_j = \{i \in \phi_r : z_{rij}^2 > n_r \tau\} ,$$

then

$$\mathcal{B} \subset \mathcal{A}.$$

Hence

$$\frac{1}{N_r} \sum_{i \in \mathcal{B}} \|z_{ri.}\|^2 \leq \frac{1}{N_r} \sum_{i \in \mathcal{A}} \|z_{ri.}\|^2 . \quad (3.C.22)$$

We have for $k=1, 2, \dots, p$, that

$$\frac{1}{N_r} \sum_{i \in \mathcal{B}_k} z_{rik}^2 \leq \frac{1}{N_r} \sum_{i \in \mathcal{B}_k} \sum_{j=1}^p z_{rij}^2 \quad (3.C.23)$$

and

$$\frac{1}{N_r} \sum_{i \in \mathcal{B}_k} \sum_{j=1}^p z_{rij}^2 \leq \frac{1}{N_r} \sum_{i \in \mathcal{B}} \sum_{j=1}^p z_{rij}^2 . \quad (3.C.24)$$

Using relations (3.C.23) and (3.C.24), we obtain

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in \mathcal{A}_k} z_{rik}^2 = 0 \quad \text{for } k=1, 2, \dots, p. \quad //$$

We next state and prove the multivariate extension to
Theorem 3.C.4.

Let

$$B_r = \frac{1}{N_r} \sum_{i=1}^{N_r} w_{ri} \cdot w_{ri}. \quad (3.C.25)$$

where

$$w_{ri} = \left(\frac{Y_{ri1} - \bar{Y}_{N_r 1}}{\sigma_{Yr1}}, \dots, \frac{Y_{rip} - \bar{Y}_{N_r p}}{\sigma_{Yrp}} \right)$$

$$\bar{Y}_{N_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} Y_{rij}$$

$$\sigma_{Yrj}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} (Y_{rij} - \bar{Y}_{N_r j})^2, \quad j=1, 2, \dots, p.$$

Assume that

$$(i) \lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{w_{rij}^2 > n_r^{-\tau}} w_{rij}^2 = 0 \quad \text{for } j=1, 2, \dots, p,$$

and that

$$(ii) \lim_{r \rightarrow \infty} B_r = B \quad (3.C.26)$$

where B_r and B are positive definite matrices. We assume B_r is positive definite for convenience.

Theorem 3.C.6

Given assumptions (i) and (ii), the limiting distribution of the p -dimensional vector $(Z_{n_r 1}, \dots, Z_{n_r p})$, where,

$$Z_{n_r j} = \frac{C_{rj}}{[V(C_{rj})]^{1/2}}, \quad j=1, 2, \dots, p, \quad (3.C.27)$$

$$C_{rj} = \sum_{i=1}^{n_r} (Y_{rij} - \bar{Y}_{N_r j})$$

$$V(C_{rj}) = \frac{n_r}{N_r} \frac{N_r - n_r}{N_r - 1} \sum_{i=1}^{N_r} (Y_{rij} - \bar{Y}_{N_r j})^2$$

is multivariate normal with mean zero and covariance matrix B .

Proof:

We consider a linear combination of the W_{rij} 's,

$$X_{ri} = \sum_{j=1}^p \lambda_j W_{rij}$$

where $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$ is any real vector.

Now,

$$\begin{aligned}
 \bar{X}_{N_r} &= \frac{1}{N_r} \sum_{i=1}^{N_r} X_{ri} \\
 &= \frac{1}{N_r} \sum_{i=1}^{N_r} \sum_{j=1}^p \lambda_j W_{rij} \\
 &= 0 .
 \end{aligned}$$

Define

$$G_{rij} = \lambda_j W_{rij} . \quad (3.C.28)$$

Since

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i=1}^{N_r} W_{rij}^2 > n_r \tau \quad W_{rij}^2 = 0 ,$$

we have that

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i=1}^{N_r} G_{rij}^2 > n_r \tau \quad G_{rij}^2 = 0 . \quad (3.C.29)$$

Using Lemma 3.C.3, equation (3.C.29) implies that

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i \in A} \| G_{ri} \|^2 = 0 \quad (3.C.30)$$

where

$$A = \{i \in \phi_r : \|G_{ri}\|^2 > n_r \tau\}$$

and

$$\|G_{ri}\|^2 = \sum_{j=1}^P G_{rij}^2.$$

Now,

$$\begin{aligned} x_{ri}^2 &= \left(\sum_{j=1}^P G_{rij} \right)^2 \\ &= \sum_{j=1}^P G_{rij}^2 + \sum_{j \neq k} G_{rij} G_{rik} \\ &\leq \sum_{j=1}^P G_{rij}^2 + \frac{1}{2} \sum_{j \neq k} (G_{rij}^2 + G_{rik}^2) \\ &= 3 \sum_{j=1}^P G_{rij}^2. \end{aligned}$$

Define the sets

$$D = \{i \in \phi_r : x_{ri}^2 \geq n_r \tau\}$$

and

$$F = \{i \in \phi_r : \sum_{j=1}^P G_{rij}^2 > n_r \frac{\tau}{3}\}.$$

Then

$$D \subset F .$$

Hence,

$$\begin{aligned} & \frac{1}{N_r} \sum_{x_{ri}^2 > \tau n_r} \left(\sum_{j=1}^P \lambda_j w_{rij} \right)^2 \\ &= \frac{1}{N_r} \sum_{i \in D} \left(\sum_{j=1}^P G_{rij} \right)^2 \\ &\leq \frac{3}{N_r} \sum_{i \in F} \sum_{j=1}^P G_{rij}^2 \\ &= \frac{3}{N_r} \sum_{i \in F} \| G_{ri.} \|^2 . \end{aligned} \tag{3.C.31}$$

Since τ is arbitrary, using condition (3.C.30) and equation (3.C.31), we obtain that

$$\lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{x_{ri}^2 > \tau n_r} x_{ri}^2 = 0 .$$

Therefore by Theorem 3.C.4,

$$\frac{\bar{X}_{n_r} - \bar{X}_{N_r}}{[V(\bar{X}_{n_r} - \bar{X}_{N_r})]^{1/2}} \xrightarrow{L} N(0,1) , \tag{3.C.32}$$

where

$$\bar{X}_{n_r} = \frac{1}{n_r} \sum_{i=1}^{n_r} X_{ri}$$

and

$$\bar{X}_{N_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} X_{ri} .$$

We have the required result by Theorem 2.C.1. //

We have now presented the work of three authors on the limiting distribution of a properly normalized mean, given that this mean arises from a simple random sample selected W.O.R. from a finite population. Madow's proof used Condition W which is a condition on all the moments. Erdős-Rényi's condition may be regarded as a pseudo-Lindeberg condition on the elements of the population. They prove that the limit of the characteristic function for $F_{N_r, n_r}(t)$ is the characteristic function for the normal distribution with mean zero and variance one. A more elegant proof is introduced by Hajek using Poisson sampling. By proving the asymptotic equivalence of S.R.S. and Poisson sampling, Hajek obtains asymptotic normality directly from the Lindeberg Central Limit Theorem.

D. A Central Limit Theorem for Regression in a Finite Population Using Hajek's Theorem

We earlier discussed the Central Limit Theorem given by Hájek [1960] for simple random samples selected without replacement from a finite population and extended this Central Limit Theorem to the case where the sampled elements were p -dimensional vectors. In this section, we investigate the limiting behavior of the estimated regression coefficients, as both the sample size and the population size become large, using Hájek's theorem.

Since we are interested in the large sample properties of estimators, we examine such properties with respect to a sequence of populations.

Accordingly, let

- i) $\phi_1, \phi_2, \dots, \phi_r$ be a sequence of populations,
- ii) N_1, N_2, \dots, N_r denote the corresponding population sizes, where $N_r > N_{r-1}$,
- iii) n_1, n_2, \dots, n_r denote the corresponding simple random samples selected without replacement, where $n_r > n_{r-1}$,
- iv) f_1, f_2, \dots, f_r denote the corresponding sampling fractions where $\lim_{r \rightarrow \infty} f_r = f$, $0 < f < 1$ and $f_r < 1 - \epsilon$, $\epsilon > 0$ for each $r=1, 2, \dots$.

The $(p+1)$ dimensional vector $(Y_{ri1}, Y_{ri2}, \dots, Y_{ri, p+1})$ is associated with the i^{th} element of the r^{th} population. We normalize this vector and define the normalized vector as

$$z_{ri.} = (z_{ri1}, z_{ri2}, \dots, z_{rip}, z_{r, i, p+1}) \quad (3.D.1)$$

where

$$z_{rij} = \frac{y_{rij} - \bar{y}_{N_r j}}{\sigma_{y_{rj}}}$$

$$\bar{y}_{N_r j} = \frac{1}{N_r} \sum_{i=1}^{N_r} y_{rij} \quad j=1, 2, \dots, p+1$$

$$\sigma_{y_{rj}}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} (y_{rij} - \bar{y}_{N_r j})^2. \quad (3.D.2)$$

Let

$$A_r = \begin{pmatrix} Q_{N_r} & H_{N_r} \\ H_{N_r}' & 1 \end{pmatrix} = \frac{1}{N_r} \sum_{i=1}^{N_r} z_{ri.}' z_{ri.},$$

where Q_{N_r} is a $(p \times p)$ matrix, H_{N_r} is a $(p \times 1)$ vector and F_{N_r} is a scalar. We define the population regression vector for the r^{th} population as

$$B_r = Q_{N_r}^{-1} H_{N_r}. \quad (3.D.3)$$

We now proceed to define the sample regression vector, but first let

$$z_{ri.} = (z_{ri1}, z_{ri2}, \dots, z_{r,i,p+1}) \quad (3.D.4)$$

where

$$z_{rij} = \frac{Y_{rij} - \bar{Y}_{n_r j}}{s_{Yrj}}$$

$$\bar{Y}_{n_r j} = \frac{1}{n_r} \sum_{i=1}^{n_r} Y_{rij} \quad j=1, 2, \dots, p+1$$

$$s_{Yrj}^2 = \frac{1}{n_r} \sum_{i=1}^{n_r} (Y_{rij} - \bar{Y}_{n_r j})^2. \quad (3.D.5)$$

Let

$$a_r = \begin{pmatrix} Q_{n_r} & H_{n_r} \\ H'_{n_r} & 1 \end{pmatrix} = \frac{1}{n_r} \sum_{i=1}^{n_r} z'_{ri.} z_{ri.},$$

where Q_{n_r} is a $(p \times p)$ matrix, H_{n_r} is a $(p \times 1)$ vector and F_{n_r} is a scalar. We define the sample regression vector for the r^{th} population as

$$b_r = Q_{n_r}^{-1} H_{n_r}. \quad (3.D.6)$$

Before proceeding to the main theorem of this section, we give the following preliminary theorem.

Theorem 3.D.1

Let $\{X_{N_r}\}$ be a sequence of independent r.v.'s with d.f.'s $\{F_{N_r}\}$ with finite means $E\{X_{N_r}\}$; and

$$S_{N_r} = \sum_{j=1}^{N_r} X_j .$$

Let b_{N_r} be a given sequence of real numbers increasing to $+\infty$.

Suppose that for some $\delta > 0$,

$$i) \frac{1}{b_{N_r}^{1+\delta}} \sum_{j=1}^{N_r} \int_{-\infty}^{\infty} |x|^{1+\delta} dF_j(x) = o(1) ;$$

then

$$\frac{1}{b_{N_r}} (S_{N_r} - a_{N_r}) \xrightarrow{P} 0$$

where

$$a_{N_r} = \sum_{j=1}^{N_r} \int_{-\infty}^{\infty} x dF_j(x) .$$

Proof:

Define for each $N_r \geq 1$ and $1 \leq j \leq N_r$:

$$\begin{aligned} \hat{X}_j &= X_j , & \text{if } |X_j| \leq b_{N_r} \\ &= 0 , & \text{otherwise.} \end{aligned}$$

Using condition i)

$$\begin{aligned}
 \sum_{j=1}^{N_r} P(X_j \neq \hat{X}_j) &= \sum_{j=1}^{N_r} P(X_j > b_{N_r}) \\
 &= \sum_{j=1}^{N_r} \int_{|x| > b_{N_r}} dF_j(x) \\
 &\leq \sum_{j=1}^{N_r} \frac{1}{b_{N_r}^{1+\delta}} \int_{|x| > b_{N_r}} |x|^{1+\delta} dF_j(x) \\
 &\leq \sum_{j=1}^{N_r} \frac{1}{b_{N_r}^{1+\delta}} \int_{-\infty}^{\infty} |x|^{1+\delta} dF_j(x) \\
 &= o(1) .
 \end{aligned}$$

Therefore, X_j and \hat{X}_j are convergent equivalent sequences by Theorem 5.2.1 in Chung [1968, p. 101]. Let

$$\tilde{X}_j = \hat{X}_j - E(\hat{X}_j) .$$

Now the \tilde{X}_j are pairwise independent by Theorem 3.3.1 in Chung [1968, p. 48]. They are also uncorrelated, since each, being bounded, has a finite second moment. Let us calculate $\sigma^2(\tilde{S}_{N_r})$ where

$$\tilde{S}_{N_r} = \sum_{j=1}^{N_r} \tilde{X}_j .$$

$$\begin{aligned}
\sigma^2(\tilde{S}_{N_r}) &= \sum_{j=1}^{N_r} \sigma^2(\tilde{X}_j) \\
&\leq \sum_{j=1}^{N_r} E(\tilde{X}_j^2) \\
&= \sum_{j=1}^{N_r} \int_{|x| \leq b_{N_r}} x^2 dF_j(x) .
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\sigma^2(\tilde{S}_{N_r})}{b_{N_r}^2} &\leq \frac{1}{b_{N_r}^2} \sum_{j=1}^{N_r} \int_{|x| \leq b_{N_r}} x^2 dF_j(x) \\
&\leq \frac{1}{b_{N_r}^{1+\delta}} \sum_{j=1}^{N_r} \int_{|x| \leq b_{N_r}} |x|^{1+\delta} dF_j(x) \\
&\leq \frac{1}{b_{N_r}^{1+\delta}} \sum_{j=1}^{N_r} \int_{-\infty}^{\infty} |x|^{1+\delta} dF_j(x) \\
&= o(1) .
\end{aligned}$$

But $\sigma^2(\tilde{S}_{N_r}) = \sigma^2(\hat{S}_{N_r})$, hence we obtain

$$\frac{\hat{S}_{N_r} - a_{N_r}}{b_{N_r}} = o_p(1) .$$

Since X_{N_r} and \hat{X}_{N_r} are convergent equivalent,

$$\frac{S_{N_r} - a_{N_r}}{b_{n_r}} = o_p(1) . \quad //$$

Recall that in Section C of Chapter III, we introduced Poisson sampling to obtain sums of independent random variables. Let the size of the Poisson sample be k_r . Define

$$t_{Yrj}^2 = \frac{1}{n_r} \sum_{i=1}^{n_r} (Y_{rij} - \bar{Y}_{N_rj})^2 \quad (3.D.7)$$

and

$$t_{Yrj}^{*2} = \frac{1}{n_r} \sum_{i=1}^{k_r} (Y_{rij} - \bar{Y}_{N_rj})^2 . \quad (3.D.8)$$

The following lemma demonstrates that as $n_r \rightarrow \infty$ and $N_r - n_r \rightarrow \infty$, the limiting variances and distributions of the random variables

t_{Yrj}^2 and t_{Yrj}^{*2} are the same.

Lemma 3.D.1

The following inequality

$$\frac{E(t_{Yrj}^2 - t_{Yrj}^{*2})^2}{V(t_{Yrj}^{*2})} \leq \left(\frac{1}{n_r} + \frac{1}{N_r - n_r} \right)^{\frac{1}{2}}, \quad (3.D.9)$$

is true.

Proof:

The proof parallels that of Hájek for Lemma 3.C.1, but we give the detailed proof for completeness. Denote by S_{n_r} the set of n_r elements drawn from ϕ_r using a simple random sampling scheme w.o.r., that is

$$S_{n_r} = \{Y_{r1j}, Y_{r2j}, \dots, Y_{rn_r j}\}, \quad j=1, 2, \dots, p+1.$$

Denote by S_{k_r} the set of k_r elements drawn using Poisson sampling.

Then $S_{n_r} \cap S_{k_r}^c$ and $S_{k_r} \cap S_{n_r}^c$ represent a simple random sample of size $|k_r - n_r|$.

$$\begin{aligned} E[(t_{Yrj}^2 - t_{Yrj}^{*2})^2] \\ = E[E(t_{Yrj}^2 - t_{Yrj}^{*2})^2 | k_r]. \end{aligned} \quad (3.D.10)$$

Define

$$W_i = (Y_{rij} - \bar{Y}_{N_r j})^2. \quad (3.D.11)$$

Then

$$\begin{aligned}
 E\{(t_{Yrj}^2 - t_{Yrj}^{*2}) | k_r\} &= E\left\{\left(\frac{1}{n_r} \sum_{i=1}^{k_r} w_i - \frac{1}{n_r} \sum_{i=1}^{n_r} w_i\right)^2 | k_r\right\} \\
 &= \frac{|k_r - n_r|}{N_r} \frac{N_r - |k_r - n_r|}{N_r - 1} \frac{1}{n_r^2} \sum_{i=1}^{N_r} (w_i^2 - \bar{w})^2 \\
 &\leq \frac{|k_r - n_r|}{N_r n_r^2} \sum_{i=1}^{N_r} w_i^2, \quad (3.D.12)
 \end{aligned}$$

where

$$\bar{w} = \frac{1}{N_r} \sum_{i=1}^{N_r} w_i.$$

Now

$$E|k_r - n_r| \leq \sqrt{E(k_r - n_r)^2} = \left(n_r \left(1 - \frac{n_r}{N_r}\right)\right)^{\frac{1}{2}}. \quad (3.D.13)$$

Using (3.D.12) and (3.D.13) we obtain

$$E(t_{Yrj}^2 - t_{Yrj}^{*2})^2 \leq \frac{1}{n_r^2} \sqrt{\left(1 - \frac{n_r}{N_r}\right) n_r} \cdot \frac{1}{N_r} \sum_{i=1}^{N_r} w_i^2. \quad (3.D.14)$$

After some algebraic manipulation we obtain that

$$V(t_{Yrj}^{*2}) = \left(\frac{1}{n_r N_r} - \frac{1}{N_r^2} \right) \sum_{i=1}^{N_r} w_i^2. \quad (3.D.15)$$

Dividing (3.D.14) by (3.D.15), we obtain the required result. //

Define

$$t_{Yr1Yrm} = \frac{1}{n_r} \sum_{i=1}^{n_r} (Y_{ril} - \bar{Y}_{N_r 1})(Y_{rim} - \bar{Y}_{N_r m})$$

and

$$t_{Yr1Yrm}^* = \frac{1}{n_r} \sum_{i=1}^{k_r} (Y_{ril} - \bar{Y}_{N_r 1})(Y_{rim} - \bar{Y}_{N_r m}).$$

Then following the proof of Lemma 3.D.1 we obtain

Corollary 3.D.1

The following inequality

$$\frac{E(t_{Yr1Yrm} - t_{Yr1Yrm}^*)^2}{V(t_{Yr1Yrm}^*)} \leq \sqrt{\frac{1}{n_r} + \frac{1}{N_r - n_r}}$$

is true.

Using Corollary 3.D.1 and Lemma 3.D.1, if the limiting distributions and variances of t_{Yr1Yrm} and t_{Yr1Yrm}^* , t_{Yr1}^2 , and t_{Yrm}^{*2} exist

under the same conditions, then they have the same limiting distributions and variances. The random variables t_{Yr1}^{*2} and t_{Yr1Yrm}^* are sums of independent random variables. We may therefore apply the well-known theory of summation of independent random variables to t_{Yr1}^{*2} and t_{Yr1Yrm}^* .

Lemma 3.D.2

Assume that for some $\delta > 0$,

$$\lim_{N_r \rightarrow \infty} \frac{1}{N_r} \sum_{i=1}^{N_r} \left(\frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Yrj}} \right)^{2+2\delta} = A < \infty, \quad (3.D.16)$$

$$j=1, 2, \dots, p+1.$$

Define

$$\rho_{r1j} = \frac{1}{N_r} \sum_{i=1}^{N_r} \left(\frac{Y_{ri1} - \bar{Y}_{N_r 1}}{\sigma_{Yr1}} \right) \left(\frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Yrj}} \right)$$

and

$$\hat{\rho}_{r1j} = \frac{t_{Yr1Yrj}}{\sigma_{Yr1} \sigma_{Yrj}} \quad 1, j=1, 2, \dots, p+1 \text{ for } 1 \neq j.$$

Then

$$\frac{t_{Yrj}^2}{\sigma_{Yrj}^2} = 1 + o_p(1) \quad (3.D.17)$$

and

$$\hat{\rho}_{rlj} = \rho_{rlj} + o_p(1) . \quad (3.D.18)$$

Proof:

We first prove

$$\frac{t_{Yrj}^{*2}}{\sigma_{Yrj}^2} = 1 + o_p(1) , \quad (3.D.19)$$

where

$$\frac{t_{Yrj}^{*2}}{\sigma_{Yrj}^2} = \frac{1}{n_r} \sum_{i=1}^{k_r} \left(\frac{Y_{rij} - \bar{Y}_{N_rj}}{\sigma_{Yrj}} \right)^2 .$$

Now

$$n_r \frac{t_{Yrj}^{*2}}{\sigma_{Yrj}^2}$$

is the sum of N_r independent random variables. We denote this sum by S_{N_rj} ,

$$S_{N_rj} = n_r \frac{t_{Yrj}^{*2}}{\sigma_{Yrj}^2} = \sum_{i=1}^{N_r} \zeta_{rij}^2 , \quad (3.D.20)$$

where

$$P[\zeta_{rij} = \frac{Y_{rij} - \bar{Y}_{N_{rj}}}{\sigma_{Y_{rj}}} = \frac{n_r}{N_r}]$$

and

$$P[\zeta_{rij} = 0] = 1 - \frac{n_r}{N_r}. \quad (3.D.21)$$

Let $F_i(x)$ denote the distribution function of ζ_{rij}^2 . The expected value of $S_{N_{rj}}$ is $a_{N_{rj}}$ where

$$\begin{aligned} a_{N_{rj}} &= \sum_{i=1}^{N_r} \int_{-\infty}^{\infty} x dF_i(x) \\ &= \sum_{i=1}^{N_r} \zeta_{rij}^2 \frac{n_r}{N_r} \\ &= n_r. \end{aligned}$$

We verify condition i) of Theorem 3.D.1 with $b_{N_r} = n_r$;

$$\frac{1}{n_r^{1+\delta}} \sum_{i=1}^{N_r} \int_{-\infty}^{\infty} |x|^{1+\delta} dF_i(x)$$

$$\begin{aligned}
&= \frac{1}{N_r^{1+\delta} f_r^{1+\delta}} \sum_{i=1}^{N_r} \left(\frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Yrj}} \right)^{2+2\delta} \frac{n_r}{N_r} \\
&= \frac{1}{\delta} \frac{1}{N_r} \sum_{i=1}^{N_r} \left(\frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Yrj}} \right)^{2+2\delta} \quad (3.D.22) \\
&= o(1) .
\end{aligned}$$

Since condition i) of Theorem 3.D.1 is satisfied, equation (3.D.17) holds. To see that relation (3.D.18) holds, we observe that

$$\begin{aligned}
&\frac{1}{N_r} \sum_{i=1}^{N_r} \left| \frac{Y_{ril} - \bar{Y}_{N_r l}}{\sigma_{Yrl}} \right|^{1+\delta} \left| \frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Yrj}} \right|^{1+\delta} \\
&< \frac{1}{2N_r} \left[\sum_{i=1}^{N_r} \left(\frac{Y_{ril} - \bar{Y}_{N_r l}}{\sigma_{Yrl}} \right)^{2+2\delta} + \sum_{i=1}^{N_r} \left(\frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Yrj}} \right)^{2+2\delta} \right] \\
&< \infty . \quad (3.D.23)
\end{aligned}$$

Now, $E(\hat{\rho}_{rlj}) = \rho_{rlj}$. Using relation (3.D.23) and Theorem 3.D.1, relation (3.D.18) immediately follows. //

We now state and prove the main theorem of this section. The following assumptions will be utilized

$$i) \quad \frac{1}{N_r^{1+\delta}} \sum_{i=1}^{N_r} Z_{rij}^{2+2\delta} = o(1) \text{ for some } \delta > 0 \text{ and } j=1, 2, \dots, p.$$

ii) The matrices A_r , a_r and A are positive definite where

$$A = \lim_{r \rightarrow \infty} A_r = \begin{pmatrix} Q & H \\ H' & 1 \end{pmatrix},$$

$$Q = \lim_{r \rightarrow \infty} Q_{N_r}$$

$$H = \lim_{r \rightarrow \infty} H_{N_r}.$$

$$\text{iii) } \lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{d_{rij}^2 > \tau n_r} d_{rij}^2 = 0$$

where

$$d_{rij} = z_{rij} \frac{e_{ri}}{\sigma_{re}}$$

$$e_{ri} = z_{r,i,p+1} - \sum_{j=1}^p B_{rj} z_{rij}$$

$$\sigma_{re}^2 = \frac{1}{N_r} \sum_{i=1}^{N_r} e_{ri}^2$$

$$B_r' = (B_{r1}, B_{r2}, \dots, B_{rp}).$$

$$\text{iv) } \lim_{r \rightarrow \infty} \frac{1}{N_r} \sum_{i=1}^{N_r} d'_{ri} \cdot d_{ri} = G$$

where

$$d_{ri} = (d_{ri1}, d_{ri2}, \dots, d_{rip}) .$$

Theorem 3.D.2

Given assumptions i) through iv),

$$\left(\frac{1}{n_r} - \frac{1}{N_r} \right)^{-\frac{1}{2}} \left(\frac{b_r - B_r}{\sigma_{re}} \right) \xrightarrow{L} N(0, Q^{-1} G Q^{-1}),$$

where b_r was defined in (3.D.6), B_r was defined in (3.D.3), Q was defined in assumption ii) and G was defined in assumption iv).

Proof:

We prove that $a_r = A_r + o_p(1)$. The lm^{th} element of $a_r - A_r$ is

$$\begin{aligned} & \frac{1}{n_r} \sum_{i=1}^{n_r} z_{ril} z_{rim} - \frac{1}{N_r} \sum_{i=1}^{N_r} z_{ril} z_{rim} \\ &= \frac{1}{n_r} \sum_{i=1}^{n_r} (z_{ril} z_{rim}) \frac{\sigma_{Yrl} \sigma_{Yrm}}{s_{Yrl} s_{Yrm}} - \left(\frac{\bar{Y}_{n_r l} - \bar{Y}_{N_r l}}{s_{Yrl}} \right) \left(\frac{\bar{Y}_{n_r m} - \bar{Y}_{N_r m}}{s_{Yrm}} \right) \\ & \quad - \frac{1}{N_r} \sum_{i=1}^{N_r} z_{ril} z_{rim} . \end{aligned} \tag{3.D.23}$$

As the result is not immediately obvious from (3.D.23), we prove that

$$\frac{s_{Yrj}^2}{\sigma_{Yrj}^2} = 1 + o_p(1), \quad j=1, 2, \dots, p+1.$$

Now,

$$\begin{aligned} \frac{s_{Yrj}^2}{\sigma_{Yrj}^2} &= \frac{1}{n_r} \sum_{i=1}^{n_r} z_{rij}^2 - \frac{(\bar{Y}_{n_rj} - \bar{Y}_{N_rj})^2}{\sigma_{Yrj}^2} \\ &= \frac{t_{Yrj}^2}{\sigma_{Yrj}^2} - \left(\frac{\bar{Y}_{n_rj} - \bar{Y}_{N_rj}}{\sigma_{Yrj}} \right)^2. \end{aligned} \quad (3.D.24)$$

Using Lemma 3.D.2,

$$\frac{t_{Yrj}^2}{\sigma_{Yrj}^2} = 1 + o_p(1). \quad (3.D.25)$$

Using the Tchebychev inequality

$$P \left\{ \left| \frac{\bar{Y}_{n_rj} - \bar{Y}_{N_rj}}{\sigma_{Yrj}} \right| \geq \epsilon \right\} \leq (1 - f_r) \frac{1}{n_r \epsilon^2} \quad \text{for } j=1, 2, \dots, p+1,$$

and Theorem 2.A.4 we obtain

$$\frac{\bar{Y}_{n_r j} - \bar{Y}_{N_r j}}{\sigma_{Yrj}}^2 = o_p(1). \quad (3.D.26)$$

Substituting equations (3.D.25) and (3.D.26) into equation (3.D.24),

$$\frac{s_{Yrj}^2}{\sigma_{Yrj}^2} = 1 + o_p(1). \quad (3.D.27)$$

Substituting equation (3.D.27) into equation (3.D.23), the l_m^{th} element of $(a_r - A_r)$ is

$$\frac{1}{n_r} \sum_{i=1}^{n_r} z_{ril} z_{rim} - \frac{1}{N_r} \sum_{i=1}^{N_r} z_{ril} z_{rim} + o_p(1). \quad (3.D.28)$$

Using Lemma 3.D.2 on equation (3.D.28) we have

$$a_r = A_r + o_p(1). \quad (3.D.29)$$

The sample regression vector is

$$b_r = Q_{n_r}^{-1} H_{n_r}. \quad (3.D.30)$$

The l^{th} element of H_{n_r} is

$$\frac{1}{n_r} \sum_{i=1}^{n_r} z_{ril}^2 z_{r,i,p+1}$$

$$\begin{aligned}
&= \frac{1}{n_r} \sum_{i=1}^{n_r} z_{ril} z_{r,i,p+1} - \left(\frac{\bar{X}_{n_r,1} - \bar{X}_{N_r,1}}{s_{Yr1}} \right) \left(\frac{\bar{X}_{n_r,p+1} - \bar{X}_{N_r,p+1}}{s_{Y,r,p+1}} \right) \\
&= \frac{1}{n_r} \sum_{i=1}^{n_r} z_{ril} z_{r,i,p+1} + o_p(1) . \tag{3.D.31}
\end{aligned}$$

Substituting equation (3.D.31) into equation (3.D.30), we obtain

$$b_r = B_r + Q_{n_r}^{-1} \frac{1}{n_r} \sum_{i=1}^{n_r} d_{ri} \cdot \sigma_{re} + o_p(1) \tag{3.D.32}$$

where d_{ri} was defined in condition iv). Now, using equation (3.D.29), $Q_{n_r} = Q_{N_r} + o_p(1)$. Since the inverses of Q_{n_r} and of Q_{N_r} exist and are continuous functions of Q_{n_r} and Q_{N_r} , utilizing a Taylor expansion (c.f. Corollary 2.A.1) we obtain

$$Q_{n_r}^{-1} = Q_{N_r}^{-1} + o_p(1).$$

Hence

$$b_r = B_r + Q_{N_r}^{-1} \frac{1}{n_r} \sum_{i=1}^{n_r} d_{ri} \cdot \sigma_{re} + o_p(1) . \tag{3.D.33}$$

The expected value of

$$\frac{1}{n_r} \sum_{i=1}^{n_r} d_{ri} \cdot \sigma_{re}$$

is zero and its variance is

$$\left(\frac{1}{n_r} - \frac{1}{N_r}\right) \frac{1}{N_r - 1} \sum_{i=1}^{N_r} d'_{ri} \cdot d_{ri} \cdot$$

Using equation (3.D.29) and Theorem 2.B.1,

$$\left(\frac{1}{n_r} - \frac{1}{N_r}\right)^{-\frac{1}{2}} \frac{b_r - B_r}{\sigma_{re}} = \left(\frac{1}{n_r} - \frac{1}{N_r}\right)^{-\frac{1}{2}} Q_{N_r}^{-1} \frac{1}{n_r} \sum_{i=1}^{n_r} d'_{ri} + o_p(1) \cdot$$

Thus, utilizing Theorem 3.D.6,

$$\left(\frac{1}{n_r} - \frac{1}{N_r}\right)^{-\frac{1}{2}} \sum_{i=1}^{n_r} \frac{d'_{ri}}{n_r} \xrightarrow{L} N(0, G) \cdot$$

The required result is obtained by using assumption ii). //

E. A Superpopulation Approach to the Limiting Distribution of Regression Coefficients in a Finite Population

In Chapter I, we saw that Cochran [1939] conducted an investigation where the finite population was viewed as a sample from an infinite population. In this section, we adopt the same approach to obtain the limiting distribution of regression coefficients computed from a sample selected from a finite population. We assume that the multivariate $(p+1)$ vector $(y, x_1, x_2, \dots, x_p)$ has mean $(\mu_y, \mu_{x_1}, \dots, \mu_{x_p})$ and nonsingular covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{yy} & \mathbf{\Sigma}_{xy} \\ \mathbf{\Sigma}_{yx} & \mathbf{\Sigma}_{xx} \end{pmatrix} \quad (3.E.1)$$

where $\mathbf{\Sigma}_{xy}$ is a $(p \times 1)$ vector and $\mathbf{\Sigma}_{xx}$ is a $(p \times p)$ matrix. We define the infinite population vector to be

$$\boldsymbol{\beta} = \mathbf{\Sigma}_{xx}^{-1} \mathbf{\Sigma}_{xy} . \quad (3.E.2)$$

We assume a sequence of finite populations $\phi_1, \phi_2, \dots, \phi_r, \dots$, with the properties i) - iv) described at the beginning of Section D in Chapter III, is drawn from the infinite population. Let

$$(y_{ri}, x_{ri}.), \quad i=1, 2, \dots, N_r$$

where

$$x_{ri}. = (x_{ri1}, x_{ri2}, \dots, x_{rip}),$$

be the i^{th} selected vector in the r^{th} population. We define the r^{th} population regression vector as

$$\mathbf{B}_r = \left(\sum_{i=1}^{N_r} x'_{ri}. x_{ri}. \right)^{-1} \left(\sum_{i=1}^{N_r} x'_{ri}. y_{ri} \right), \quad (3.E.3)$$

and the associated r^{th} sample regression vector as

$$\mathbf{b}_r = \left(\sum_{i=1}^{n_r} x'_{ri}. x_{ri}. \right)^{-1} \left(\sum_{i=1}^{n_r} x'_{ri}. y_{ri} \right), \quad (3.E.4)$$

where N_r is the size of the r^{th} population sampled and n_r is the size of the simple random sample selected w.o.r. from the finite population. Let the infinite population residual be

$$e = y - x_{\cdot} \beta, \quad (3.E.5)$$

where

$$x_{\cdot} = (x_1, x_2, \dots, x_p) \cdot$$

Define

$$\theta = E\{(x_{\cdot} e)' (x_{\cdot} e)\}, \quad (3.E.7)$$

θ being a $(p \times p)$ symmetric matrix. We define the $(n_r \times p)$ sample matrix

$$x_{n_r} = (x'_{r1\cdot}, x'_{r2\cdot}, \dots, x'_{rn_r\cdot})'$$

and the $(N_r \times p)$ population matrix

$$x_{N_r} = (x'_{r1\cdot}, x'_{r2\cdot}, \dots, x'_{rN_r\cdot})'.$$

We give a theorem closely related to Theorem 1 stated in Fuller [1973].

Theorem 3.E.1

Let a sequence of simple random samples of size n_r be selected from a sequence of finite populations ϕ_r of size N_r . Let this finite population be a random sample from an infinite population with finite fourth moments. Let

$$f_r = \frac{n_r}{N_r}$$

where $0 < f_r < 1$, $\lim_{r \rightarrow \infty} f_r = f$. Then as $r \rightarrow \infty$

$$n_r^{1/2} (b_r - B_r) \xrightarrow{L} N(0, (1-f) \Sigma_{xx}^{-1} \theta \Sigma_{xx}^{-1}) .$$

Proof:

By Theorem 5.2.2. in Chung [1968, p. 103],

$$\sum_{i=1}^{N_r} \frac{x'_{ri} \cdot x_{ri}}{N_r} = \Sigma_{xx} + o_p(1) . \quad (3.E.7)$$

Let e_{N_r} be an $(N_r \times 1)$ vector where the i^{th} element of e_{N_r} is

$$y_{ri} - x_{ri} \cdot \beta . \quad (3.E.8)$$

The vector e_{n_r} is the $(n_r \times 1)$ analogously defined vector associated with the sample. In view of equation (3.E.8) we have,

$$b_r - B_r = (x'_{n_r} x_{n_r})^{-1} x'_{n_r} e_{n_r} - (x'_{N_r} x_{N_r})^{-1} x'_{N_r} e_{N_r} . \quad (3.E.9)$$

Since the elements of

$$\frac{x'_{n_r} x_{n_r}}{n_r}$$

are sample moments with variances of order $\frac{1}{n_r}$, we have

$$\frac{x'_{n_r} x_{n_r}}{n_r} = \hat{\Sigma}_{xx} + O_p(n_r^{-1/2})$$

$$\frac{x'_{N_r} x_{N_r}}{N_r} = \hat{\Sigma}_{xx} + O_p(N_r^{-1/2})$$

and

$$n_r^{1/2} (b_r - B_r) = n_r^{1/2} \hat{\Sigma}_{xx}^{-1} \left(\frac{x'_{n_r} e_{n_r}}{n_r} - \frac{x'_{N_r} e_{N_r}}{N_r} \right) + O_p(n_r^{-1/2}). \quad (3.E.11)$$

Partitioning $x'_{N_r} e_{N_r}$ into

$$\begin{pmatrix} x'_{n_r} & x'_{N_r - n_r} \end{pmatrix} \begin{pmatrix} e_{n_r} \\ e_{N_r - n_r} \end{pmatrix},$$

we obtain

$$x'_{n_r} e_{n_r} - f_r x'_{N_r} e_{N_r} = (1-f_r) x'_{n_r} e_{n_r} - f_r x'_{N_r - n_r} e_{N_r - n_r}. \quad (3.E.12)$$

Now

$$x'_{n_r} e_{n_r} = \left[\sum_{i=1}^{n_r} x_{ri1} e_{ri}, \sum_{i=1}^{n_r} x_{ri2} e_{ri}, \dots, \sum_{i=1}^{n_r} x_{rip} e_{ri} \right]'$$

and

$$x'_{N_r - n_r} e_{N_r - n_r} = \left[\sum_{i=n_r+1}^{N_r} x_{ri1} e_{ri}, \sum_{i=n_r+1}^{N_r} x_{ri2} e_{ri}, \dots, \sum_{i=n_r+1}^{N_r} x_{rip} e_{ri} \right]'$$

Now $E\{x_{rij} e_{ri}\} = 0$, $j=1, 2, \dots, p$ and $E\{(x_{rij} e_{ri})^2\} < \infty$ for $j=1, 2, \dots, p$. Furthermore, the vectors $(x_{ri1} e_{ri}, x_{ri2} e_{ri}, \dots, x_{rip} e_{ri})$, $i=1, 2, \dots$ are independently and identically distributed.

Let λ be an arbitrary nonzero $(p \times 1)$ vector, and consider

$$\begin{aligned} S_{r1} &= n_r^{-1/2} \lambda' (1-f_r) x'_{n_r} e_{n_r} \\ &= n_r^{-1/2} \sum_{j=1}^p \lambda_j (1-f_r) \sum_{i=1}^{n_r} x_{rij} e_{ri} \end{aligned}$$

First, note that

$$\begin{aligned} E\{S_{r1}\} &= n_r^{-1/2} \lambda' (1-f_r) E\{x'_{n_r} e_{n_r}\} \\ &= 0 \end{aligned}$$

for all n_r . Second, note that

$$\begin{aligned} E\{(S_{r1})^2\} &= \frac{1}{n_r}(1-f_r)^2 \lambda' E\{(x'_{n_r} e_{n_r})(x'_{n_r} e_{n_r})'\} \lambda \\ &= (1-f_r)^2 \lambda' \theta \lambda . \end{aligned}$$

In a similar manner,

$$S_{r2} = (N_r - n_r)^{-\frac{1}{2}} \lambda' f_r^{\frac{1}{2}} (1-f_r)^{\frac{1}{2}} x'_{N_r - n_r} e_{N_r - n_r}$$

has mean zero and variance $f_r(1-f_r)\lambda' \theta \lambda$. Hence, by the Lindeberg-Levy Central Limit Theorem [Rao, 1965, p. 108],

$$S_{r1} \xrightarrow{L} N(0, (1-f)^2 \lambda' \theta \lambda)$$

and

$$S_{r2} \xrightarrow{L} N(0, f(1-f) \lambda' \theta \lambda) .$$

Since S_{r1} and S_{r2} are independent, it follows that

$$S_{r1} + S_{r2} \xrightarrow{L} N(0, (1-f) \lambda' \theta \lambda) .$$

The required result follows by the multivariate central limit theorem (Theorem 2.C.1). //

Fuller [1973] noted that a consistent estimator of the variance of $n_r^{1/2} (b_r - B_r)$ is easily constructed by estimating the matrix θ by

$$\hat{\theta}_r = \frac{1}{n_r - p} \sum_{i=1}^{n_r} \hat{d}_{ri}' \hat{d}_{ri}.$$

where

$$\hat{d}_{ri} = x_{ri}' \hat{e}_{ri}$$

$$\hat{e}_{ri} = y_{ri} - x_{ri}' b_r$$

$$x_{ri} = (x_{ri1}, x_{ri2}, \dots, x_{rip})$$

is the i^{th} row of the $(n_r \times p)$ matrix x_{n_r} .

Theorem 3.E.2 (Fuller)

Given the stated assumptions, a consistent estimator for $(b_r - B_r)$ is

$$\left(\frac{x_{n_r}' x_{n_r}}{n_r} \right)^{-1} \left(\frac{\hat{\theta}_r}{n_r} \right) \left(\frac{x_{n_r}' x_{n_r}}{n_r} \right)^{-1}$$

where

$$\hat{\theta}_r = (1 - f_r) \frac{1}{n_r} \sum_{i=1}^{n_r} \hat{d}_{ri}' \hat{d}_{ri}.$$

$$\hat{d}_{ri} = x_{ri}' \hat{e}_{ri}$$

$$\hat{e}_{ri} = y_{ri} - x_{ri}' b_r$$

Proof:

The lm^{th} element of $\hat{\theta}_r$ is

$$(1-f_r) \frac{1}{n_r} \sum_{i=1}^{n_r} \hat{d}_{ril} \hat{d}_{rim}$$

where

$$\hat{d}_{rim} = x_{rim} \hat{e}_{ri}, \quad m=1, 2, \dots, p.$$

$$\frac{1}{n_r} \sum_{i=1}^{n_r} \hat{d}_{ril} \hat{d}_{rim} = \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ril} [e_{ri} - x_{ri}'(b_r - \beta)]$$

$$= \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ril} x_{rim} [e_{ri} - x_{ri}'(b_r - \beta)]$$

$$= \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ril} x_{rim} e_{ri}^2 - \frac{1}{n_r} \sum_{i=1}^{n_r} e_{ri} x_{ril} x_{rim} (b_r - \beta)$$

$$- \frac{1}{n_r} \sum_{i=1}^{n_r} e_{ri} x_{rim} x_{ril} (b_r - \beta) + (b_r - \beta)' T_r (b_r - \beta)$$

where the hq^{th} element of T_r given by

$$\frac{1}{n_r} \sum_{i=1}^{n_r} x_{ril} x_{rih} x_{rim} x_{riq} = O(1)$$

by the moment assumptions. Hence,

$$\frac{1}{n_r} \sum_{i=1}^{n_r} \hat{d}_{ril} \hat{d}_{rim} = \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ril} x_{rim} e_{ri}^2 + O_p(n_r^{-1/2}) .$$

By the Weak Law of Large Numbers,

$$\frac{1}{n_r} \sum_{i=1}^{n_r} x_{ril} x_{rim} e_{ri}^2 \xrightarrow{P} E\{x_l x_m e^2\}$$

and the result follows. //

IV. ERRORS-IN-VARIABLES

A. Introduction

The study of errors-in-variables dates from the 19th century. Literature reviews devoted to this topic are contained in Kendall and Stuart [1961], Madansky [1959], Cochran [1968], Malinvaud [1966], Moran [1971], and Wolter [1974]. The problem at hand is the estimation of regression coefficients and their associated variances given that the input variables are subject to error. It has been demonstrated that data collected from human respondents, contains errors and inaccuracies. The U.S. Bureau of the Census [1972] has reported estimates of the response variance as a percent of total variance that range from 0.5 percent for items such as 5 year age classes to 40 percent for income classes.

1. The model

Let (Λ, F, P) be a probability space, with $\mathbb{H} \subset \mathbb{R}^p$, and n an integer larger than p . Let $\{z_t\}_{t=1}^{\infty}$ and $\{\epsilon_t\}_{t=1}^{\infty}$ be sequences of p -dimensional vectors; that is, a sequence of real valued functions, with argument (ω) , which are measurable (Λ, F) . We denote the elements of \mathbb{H} by θ , where θ are $(p \times 1)$ vectors.

We define a linear regression errors-in-variables model as

$$f(z_t; \theta) = z_t \theta \quad t=1, 2, \dots, \quad (4.A.1)$$

$$= 0,$$

where

$$z_t = (y_t, x_t) \in \mathbb{R}^{p+1},$$

$$\theta \in \mathbb{H} \subset \mathbb{R}^p \quad (4.A.2)$$

$$f: \mathbb{R}^{p+1} \times \mathbb{H} \longrightarrow \mathbb{R}',$$

f being a Lebesgue measurable function. Let

$$Z_t = z_t + \epsilon_t.$$

We have an errors-in-variables problem if the true value z_t can be observed only through Z_t .

The errors-in-variables model can be written in a more familiar form. Let

$$\{y_t\}_{t=1}^{\infty}, \{x_t\}_{t=1}^{\infty}, \{e_t\}_{t=1}^{\infty}, \{u_t\}_{t=1}^{\infty}, \{Y_t\}_{t=1}^{\infty}, \{X_t\}_{t=1}^{\infty}$$

be a sequence of random variables.

The errors-in-variables model is

$$y = x \beta_1, \quad (4.A.4)$$

where y is an $(n \times 1)$ vector, x is an $(n \times p)$ matrix and β_1 is a $(p \times 1)$ vector. The elements of β_1 are unknown parameters that are to be estimated. y and x can be observed through Y and X where,

$$Y = y + e,$$

$$X = x + u . \quad (4.A.5)$$

Z , z and ϵ are $(n \times (p+1))$ matrices defined as

$$Z = (Y, X) ,$$

$$z = (y, x) , \quad (4.A.6)$$

$$\epsilon = (e, u)$$

and

$$\beta = \begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix} .$$

Using (4.A.5) and (4.A.6), the errors-in-variables model can be written as in (4.A.1), namely

$$z\beta = 0 \quad (4.A.7)$$

where we observe Z , $Z = z + \epsilon$. We assume that

$$E(\epsilon_t) = 0 \quad (4.A.8)$$

for all t and

$$\Sigma_{ts} = E(\epsilon_t' \epsilon_s) = \begin{pmatrix} \sigma_{e_t e_s} & \Sigma_{e_t u_s} \\ \Sigma_{u_t e_s} & \Sigma_{u_t u_s} \end{pmatrix} . \quad (4.A.9)$$

Throughout the rest of this chapter, we will assume that

$$\Phi_{ts} = \Phi \text{ for } t = s$$

$$= 0 \text{ otherwise ,}$$

and that ϵ_t is independent of z_t for all t and s . We introduce the vector of residuals v as

$$v = \epsilon\beta = e - u\beta_1 . \quad (4.A.10)$$

If u and e in equation (4.A.5) are nonzero, Adcock [1877, 1878] suggested a procedure that has been called an "orthogonal regression". The orthogonal regression procedure minimizes the sum of squares of the perpendicular distances from the fitted line to the data points. This method's weakness is that the orthogonal regression estimator is not invariant under transformations of the coordinate system.

Kummell [1879] proposed a solution to this noninvariance by assuming that the ratio σ_u^2/σ_e^2 is known. His procedure was to minimize the sum of squares of the weighted distances from the observed points to the fitted line, where the weights are proportional to the inverses of the variances of e_t and u_t .

Koopman [1937] found the maximum likelihood estimator for the functional model $\epsilon_t \sim \text{NID}(0, \Phi)$, Φ known up to a multiple. He also derived the approximate covariance matrix of the estimator under the assumption that the structural relationship case holds, and then only when $\Phi_{xx}^{-1} (\frac{1}{n} u'u)$ is small, where Φ_{xx} is the covariance matrix

of the distribution of x .

Assuming Σ unknown, but an estimate S of Σ available, F. S. Acton [1959] suggested the estimator of β obtained by replacing Σ by S in Koopman's maximum likelihood estimator.

Villegas [1961] considered the case of unknown Σ . He obtained the maximum likelihood estimator, provided that S was distributed as a Wishart $W(\Sigma, n)$ independent of Z_t , $t=1, 2, \dots, n$.

Fuller [1971] considered several different structures of the functional relationship. He modified the maximum likelihood estimators in such a way that the existence of finite moments is guaranteed. He derived the asymptotic distribution of each estimate, and showed his estimators to have smaller mean square error than the usual maximum likelihood estimates.

B. An Errors-in-variables Regression

Model for Clusters

Fuller [1973] considered the following model for a population divided into clusters. Let

$$y_{ij} = \sum_{k=0}^p \beta_k x_{ijk}$$

$$i=1, 2, \dots, n$$

$$Y_{ij} = y_{ij} + e_{ij}$$

$$j=1, 2, \dots, m_i$$

$$X_{ijk} = x_{ijk} + u_{ijk}$$

where the index i denotes primary units and the index j secondary units.

Assume that Y_{ij} and x_{ijk} are drawn from an infinite multivariate population satisfying the following assumptions. Express the ij^{th} observation in a population of N primaries of size M_i as

$$d_{ij} = u_i + \delta_{ij} \quad i=1, 2, \dots, N,$$

$$j=1, 2, \dots, M_i.$$

It is assumed that the vectors

$$(u_i, M_i, \sigma_{\delta i}^2) \quad i=1, 2, \dots, N$$

are a random sample from a multivariate population with finite sixth moments. Assume that $E\{M_i u_i\} = 0$. Conditional upon the selection of cluster sizes it is assumed that the δ_{ij} are chosen from an infinite population with zero mean, variances $\sigma_{\delta i}^2$, and bounded third moment. Assume that the primary sampling rate, $0 < f_1 < 1$ is fixed. Let g_{2i} , $1 \geq g_{2i} > 0$ $i=1, 2, \dots$ be a fixed sequence, m_i the smallest integer greater than or equal to $g_{2i} M_i$.

Assume that $E(e_{ij} | x_{ij.}) = 0$ for all $x_{ij.}$, where $x_{ij.}$ is the vector of observations of the independent variables. Assume that the observed x_{ijk} are the sum of the unobservable true value, x_{ijk} , and an error of measurement, u_{ijk} . Assume that the error of

measurement has expected value zero for all x_{ijk} , u_{ijk} is independent of u'_{ijk} , $i \neq i'$, and the fifth moment of u_{ijk} is uniformly bounded. Let

$$E \left\{ \begin{pmatrix} y_{ij} \\ x'_{ij} \end{pmatrix} (y_{ij}, x_{ij.}) \right\} = \begin{pmatrix} \sigma_y^2 & \mathbb{E}_{yx} \\ \mathbb{E}_{xy} & \mathbb{E}_{xx} \end{pmatrix} + \begin{pmatrix} \sigma_e^2 & \mathbb{E}_{eu} \\ \mathbb{E}_{ue} & \mathbb{E}_{uu} \end{pmatrix}$$

where \mathbb{E}_{eu} and \mathbb{E}_{uu} are known, and \mathbb{E}_{xx} is nonsingular.

Theorem 4.B.1 (Fuller)

Given the stated assumptions

$$n^{\frac{1}{2}} (\hat{\beta} - \beta) \xrightarrow{L} N(0, \mathbb{E}_{xx}^{-1} G \mathbb{E}_{xx}^{-1})$$

and

$$(\hat{\mathbb{E}}_{xx} - \mathbb{E}_{uu})^{-1} \tilde{G} (\hat{\mathbb{E}}_{xx} - \mathbb{E}_{uu})^{-1}$$

is a consistent estimator for the variance of $\hat{\beta}$, where

$$\hat{\beta} = (\hat{\mathbb{E}}_{xx} - \mathbb{E}_{uu})^{-1} (\hat{\mathbb{E}}_{xy} - \mathbb{E}_{ue})$$

$$\hat{\mathbb{E}}_{xx} = \frac{1}{nM} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} x'_{ij} \cdot x_{ij}.$$

$$\hat{t}_{XY} = \frac{1}{nM} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} x'_{ij} \cdot y_{ij}$$

$$G = NE\{(\hat{t}_{xv} - t_{uv})(\hat{t}_{vx} - t_{vu})\}$$

$$\tilde{G} = \frac{1}{nM} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} (\tilde{d}_{i..} - \bar{d}_{...})' (\tilde{d}_{i..} - \bar{d}_{...})$$

$$\tilde{d}_{ij.} = x_{ij.} \tilde{v}_{ij}$$

$$\tilde{d}_{i..} = \frac{1}{m_i} \sum_{j=1}^{m_i} \tilde{d}_{ij.}$$

$$\bar{d}_{...} = \frac{1}{nM} \sum_{i=1}^n M_i \tilde{d}_{i..}$$

$$\tilde{v}_{ij} = y_{ij} - x_{ij.} \hat{\beta}$$

$$\bar{M} = \frac{1}{n} \sum_{i=1}^n M_i$$

$$\hat{t}_{xv} = \frac{1}{nM} \sum_{i=1}^n \frac{M_i}{m_i} \sum_{j=1}^{m_i} x'_{ij} \cdot v_{ij}$$

$$v_{ij} = y_{ij} - x_{ij.} \beta$$

Proof: See Fuller [1973]. //

C. Model for Clustered Data Subject to Error

1. Introduction

Fuller and Battese [1973] considered the following model for clustered data,

$$y_{it} = \sum_{l=1}^p x_{itl} \beta_l + d_{it}, \quad t=1, 2, \dots, n_i; i=1, 2, \dots, k$$

and

$$d_{it} = w_i + s_{it}$$

where y_{it} denotes the value of the t^{th} observation within the i^{th} primary; x_{itl} , $l=1, 2, \dots, p$, denotes the levels of the p control variables at which the observation y_{it} is obtained (x_{itl} is assumed to be fixed); β_l , $l=1, 2, \dots, p$, denotes the unknown parameters to be estimated; and d_{it} , the random error associated with y_{it} is assumed the sum of the random effect associated with the i^{th} primary (w_i) and the random effect associated with the t^{th} observation for the i^{th} primary (s_{it}).

The random errors w_i and s_{it} are assumed independently distributed with zero means and variances σ_w^2 and σ_s^2 , respectively. The covariance structure for the random errors d_{it} is expressed by,

$$E(d_{it}d_{i't'}) = \sigma_w^2 + \sigma_s^2, \quad \text{if } i=i', t=t'$$

$$= \sigma_w^2, \quad \text{if } i=i', t \neq t'$$

$$= 0, \quad \text{otherwise.}$$

Fuller and Battese transformed the above model from one of generalized least squares to one of simple least squares as follows.

$$y_{it} - \alpha_i \bar{y}_i = \sum_{l=1}^p (x_{itl} - \alpha_i \bar{x}_{i \cdot l}) \beta_l + d_{it}^*$$

where

$$\alpha_i = 1 - [\sigma_s^2 / (\sigma_s^2 + n_i \sigma_w^2)]^{1/2}$$

and $\bar{y}_{i \cdot}$, $\bar{x}_{i \cdot l}$, $l=1, 2, \dots, p$, denote the averages of the n_i y - and x -measurements in the i^{th} primary unit. The d_{it}^* where

$$d_{it}^* = d_{it} - \alpha_i \bar{d}_{i \cdot}$$

$$\bar{d}_{i \cdot} = \frac{1}{n_i} \sum_{t=1}^{n_i} d_{it}$$

are uncorrelated and have variances σ_s^2 . We investigate this cluster model when y and x are subject to measurement error.

2. The model and assumptions

We assume the errors-in-variables model

$$y_{it} = x_{it} \beta_1$$

$$X_{it} = x_{it} + u_{it} \quad (4.C.1)$$

$$Y_{it} = y_{it} + d_{it}$$

where

$$i=1, 2, \dots, k \quad \text{and } t=1, 2, \dots, n_i$$

β_1 is a $p \times 1$ vector of unknown parameters. The following assumptions will be utilized in analyzing model (4.C.1).

i) $\epsilon_{it} = (d_{it}; u_{it})$ is an $n_i \times (p+1)$ matrix of random variables distributed independently of x_{jm} for all i, t, j and m . For $i \neq i'$, ϵ_{it} is distributed independently of $\epsilon_{i't'}$. For each primary unit i , u_{it} is distributed independently of $u_{it'}$, $t \neq t'$.

ii) d_{it} is independent of u_{jm} for all i, j, m and t .

iii) For each primary unit i

$$d_{it} = w_i + s_{it}, \quad (4.C.2)$$

where the random errors w_i and s_{it} are independently distributed with zero means and variances σ_w^2 and σ_s^2

respectively. The covariance structure for the random errors d_{it} is expressed by,

$$\begin{aligned} E(d_{it} d_{i't'}) &= \sigma_w^2 + \sigma_s^2, & \text{if } i=i', t=t' \\ &= \sigma_w^2, & \text{if } i=i', t \neq t' \\ &= 0, & \text{otherwise.} \end{aligned}$$

- iv) $w_i \sim \text{NID}(0, \sigma_w^2)$, $s_{it} \sim \text{NID}(0, \sigma_s^2)$ and $u_{it} \sim \text{NID}(0, \sigma_{uu})$ where w_i , s_{it} and $u_{i''t'}$ are independent for all i , i' , i'' , t and t' .

Model 4.C.1 may be written as

$$\begin{aligned} Y_{it} &= y_{it} + d_{it} \\ &= x_{it} \beta_1 + d_{it} \\ &= (x_{it} - u_{it}) \beta_1 + d_{it} & (4.C.3) \\ &= x_{it} \beta_1 + (d_{it} - u_{it} \beta_1) \\ &= x_{it} \beta_1 + v_{it}, \end{aligned}$$

where

$$v_{it} = d_{it} - u_{it} \beta_1. \quad (4.C.4)$$

We write the linear model (4.C.3) as

$$Y = X \beta_1 + v \quad (4.C.5)$$

where

$$Y = (Y_1', Y_2', \dots, Y_k'), \quad Y_i' = (Y_{i1}, Y_{i2}, \dots, Y_{in_i}) ,$$

$$i=1, 2, \dots, k ,$$

$$v = (v_1', v_2', \dots, v_k')', \quad v_i' = (v_{i1}, v_{i2}, \dots, v_{in_i}) ,$$

and the X matrix is constructed similarly. We assume that the (nxp) matrix in (4.C.5) is a matrix of observed variables and has rank p, where n denotes the number of observations on the Y-variable

$$(i.e., n = \sum_{i=1}^k n_i) .$$

Now, we express the random error v_{it} associated with Y_{it} in model (4.C.3) as the sum of the random effect associated with the i^{th} sample primary (w_i) and the random effect associated with the t^{th} observation in the i^{th} primary (δ_{it}). That is,

$$\begin{aligned} v_{it} &= (w_i + s_{it}) - u_{it} \beta_1 \\ &= w_i + (s_{it} - u_{it} \beta_1) , \end{aligned}$$

and letting

$$\delta_{it} = s_{it} - u_{it} \beta_1, \quad (4.C.6)$$

write

$$v_{it} = w_i + \delta_{it}.$$

By assumption iv), w_i is independent of δ_{it} . Denote the covariance matrix for $(u_{it}; v_{it})$, $i=1, 2, \dots, k$; $t=1, 2, \dots, n_i$, as Σ_{uv} where

$$\begin{aligned} \Sigma_{uv} &= E(u'_{it} v_{it}) \\ &= E\{u'_{it} (d_{it} - u_{it} \beta_1)\} \\ &= - \Sigma_{uu} \beta_1. \end{aligned}$$

To consider the large-sample properties of this model, we consider the sequence of estimators indexed by n where

$$n = \sum_{i=1}^k n_i$$

is the sample size. We shall also need the assumption:

v) The elements of y and x , where

$$Y = y + d$$

and

$$X = x + u , \quad (4.C.7)$$

are uniformly bounded for all n . The cluster size n_i is bounded for all $i=1, 2, \dots, k$.

3. Estimation

Given the assumptions on the random errors w_i and δ_{it} , the covariance matrix of $v_i' = (v_{i1}, v_{i2}, \dots, v_{in_i})$ is an $n_i \times n_i$ block diagonal matrix A_i where

$$A_i = \sigma_\delta^2 I_{n_i} + \sigma_w^2 J_{n_i} ,$$

I_{n_i} is the identity matrix of order n_i , and J_{n_i} is the $(n_i \times n_i)$ matrix with all elements equal to one.

The characteristic roots of A_i are

$$\lambda_1 = \sigma_\delta^2 + n_i \sigma_w^2$$

$$\lambda_2 = \lambda_3 = \dots = \lambda_{n_i} = \sigma_\delta^2 ,$$

with corresponding eigenvectors

$$\begin{bmatrix} \frac{1}{\sqrt{n_i}} \\ \frac{1}{\sqrt{n_i}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{\sqrt{n_i}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{1.2}} \\ -\frac{1}{\sqrt{1.2}} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} \frac{1}{\sqrt{n_i(n_i-1)}} \\ \frac{1}{\sqrt{n_i(n_i-1)}} \\ \cdot \\ \cdot \\ \cdot \\ -\frac{n_i-1}{\sqrt{n_i(n_i-1)}} \end{bmatrix} \quad (4.C.8)$$

The n_i equations for the i^{th} primary unit can be written as

$$Y_i = X_i \beta_1 + v_i \quad i=1, 2, \dots, k \quad (4.C.9)$$

where

$$Y_i' = (Y_{i1}, Y_{i2}, \dots, Y_{in_i}) ,$$

$$v_i' = (v_{i1}, v_{i2}, \dots, v_{in_i}) ,$$

X_i is the $n_i \times p$ matrix of observed independent variables associated with the i^{th} cluster.

Transform the n_i equations for the i^{th} primary unit to

$$G_i = F_i \beta_1 + v_i \quad (4.C.10)$$

where

$$G_i = \Gamma_i' Y_i ,$$

$$F_i = \Gamma_i' X_i ,$$

$$V_i = \Gamma_i' v_i$$

and

$$\Gamma_i = \begin{bmatrix} \frac{\sigma_\delta^2}{n_i(\sigma_\delta^2 + n_i\sigma_w^2)} , & \frac{1}{\sqrt{2}} , & \dots , & \frac{1}{\sqrt{n_i(n_i-1)}} \\ \frac{\sigma_\delta^2}{n_i(\sigma_\delta^2 + n_i\sigma_w^2)} , & \frac{-1}{\sqrt{2}} , & \dots , & \frac{1}{\sqrt{n_i(n_i-1)}} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\sigma_\delta^2}{n_i(\sigma_\delta^2 + n_i\sigma_w^2)} , & 0 , & \dots , & -\frac{n_i-1}{\sqrt{n_i(n_i-1)}} \end{bmatrix} . \quad (4.C.11)$$

Lemma 4.C.1

Given that the random errors v_{ij} , $j=1, 2, \dots, n_i$; $i=1, 2, \dots, k$; have covariance structure

$$\begin{aligned}
E(v_{ij}v_{i'j'}) &= \sigma_w^2 + \sigma_\delta^2, & \text{if } i=i', j=j' \\
&= \sigma_w^2, & \text{if } i=i', j \neq j' \\
&= 0, & \text{otherwise:}
\end{aligned}$$

then the transformed errors v_{ij} where

$$v_{i1} = \frac{\sigma_\delta^2}{n_i(\sigma_\delta^2 + n_i\sigma_w^2)} \sum_{j=1}^{n_i} v_{ij};$$

and

$$v_{il} = \frac{1}{\sqrt{l(l-1)}} \left[\sum_{j=1}^{l-1} v_{ij} - (l-1) v_{il} \right], \quad l=2, 3, \dots, n_i,$$

are uncorrelated and have variance σ_δ^2 .

Proof:

The proof is by direct substitution. //

Given the model with uncorrelated measurement errors

$$Y = X\beta_1 + v,$$

$$X = x + u,$$

$$Y = y + d,$$

where

$$(d; u) \sim \text{NID} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & 0 \\ 0 & \Sigma_{uu} \end{pmatrix} \right],$$

an estimator of β is

$$\hat{\beta} = (X'X - n\Sigma_{uu})^{-1} X'Y. \quad (4.C.12)$$

Note that in such a case

$$E(X'X) = x'x + n\Sigma_{uu}.$$

In our problem the expected value of the matrix of squares and products of transformed X's is:

$$\begin{aligned} E(F'F) &= \sum_{i=1}^k E(F_i'F_i) \\ &= \sum_{i=1}^k \left\{ f_i'f_i + \left[\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i\sigma_w^2} + (n_i-1) \right] \Sigma_{uu} \right\}, \end{aligned} \quad (4.C.13)$$

where

$$\begin{aligned} F_i &= f_i + \eta_i, \\ f_i &= \Gamma_i' x_i, \\ \eta_i &= \Gamma_i' u_i. \end{aligned} \quad (4.C.14)$$

Model 4.C.4, is transformed to

$$G = F \beta_1 + V \quad (4.C.15)$$

where

$$G = g + \zeta \quad (4.C.16)$$

$$F = f + \eta \quad (4.C.17)$$

$$g = f \beta_1 \quad (4.C.18)$$

$$G = (G'_1, G'_2, \dots, G'_k)'$$

$$F = (F'_1, F'_2, \dots, F'_k)'$$

$$g = (g'_1, g'_2, \dots, g'_k)'$$

$$f = (f'_1, f'_2, \dots, f'_k)'$$

$$\zeta = (\zeta'_1, \zeta'_2, \dots, \zeta'_k)'$$

$$\eta = (\eta'_1, \eta'_2, \dots, \eta'_k)'$$

$$\beta_1 = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ip})'$$

$$G_i = \Gamma'_i Y_i, g_i = \Gamma'_i y_i, F_i = \Gamma'_i X_i, f_i = \Gamma'_i x_i,$$

$$\zeta_i = \Gamma_i' d_i \text{ and } g_i = \Gamma_i' y_i \quad \text{for } i=1, 2, \dots, k.$$

In view of equations (4.C.12) and (4.C.13), it is natural to define an estimator of β_1 for the cluster model as

$$\tilde{\beta}_1 = (F'F - n^* \Sigma_{uu})^{-1} F'G \quad (4.C.19)$$

where

$$n^* = \sum_{i=1}^k \left[\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} + (n_i - 1) \right].$$

We make the following assumption for the transformed model (4.C.5),

vi) $\frac{1}{n^*} f'f$ is a positive definite matrix for all n^* and

$$\lim_{n^* \rightarrow \infty} \frac{1}{n^*} f'f = \bar{M}_{xx}^*.$$

4. Variance of the regression coefficients

given normal error structure

In the main theorem of this section, we will utilize the following lemma, concerning the fourth moments of a multivariate normal distribution.

Lemma 4.C.2

Suppose Z_1, Z_2, \dots, Z_n are independent $(p+1)$ dimensional vectors, where $Z_i, i=1, 2, \dots, n$ is distributed according to

$$Z_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \sigma_v^2 \end{pmatrix} \right],$$

u_i being a $p \times 1$ vector and v_i a scalar. Then

$$\begin{aligned} E \left\{ \left(\frac{1}{n} \sum_{i=1}^n u_i' v_i - \Sigma_{uv} \right) \left(\frac{1}{n} \sum_{i=1}^n u_i' v_i - \Sigma_{uv} \right)' \right\} \\ = \frac{1}{n} (\sigma_v^2 \Sigma_{uu} + \Sigma_{uv} \Sigma_{vu}) . \end{aligned}$$

Proof:

Denote the $(k,1)^{th}$ element of Σ_{uu} by $\sigma_{u_k u_1}$ and the k^{th} element of Σ_{uv} by $\sigma_{u_k v}$. We proceed to find the expectation of the $(k,1)^{th}$ element of

$$\left(\frac{1}{n} \sum_{i=1}^n u_i' v_i - \Sigma_{uv} \right) \left(\frac{1}{n} \sum_{i=1}^n u_i' v_i - \Sigma_{uv} \right)' ,$$

where this element is:

$$a_{kl} = \left(\frac{1}{n} \sum_{i=1}^n u_{ik} v_i - \sigma_{u_k v} \right) \left(\frac{1}{n} \sum_{i=1}^n u_{il} v_i - \sigma_{u_l v} \right) .$$

Now,

$$\begin{aligned}
 E(a_{kl}) &= E\left\{\left(\frac{1}{n} \sum_{i=1}^n u_{ik} v_i - \sigma_{u_k v}\right) \left(\frac{1}{n} \sum_{i=1}^n u_{il} v_i - \sigma_{u_l v}\right)\right\} \\
 &= E\left\{\frac{1}{n^2} \sum_{i=1}^n u_{ik} u_{il} v_i^2\right\} + E\left\{\frac{1}{n^2} \sum_{i \neq j} u_{ik} u_{jl} v_i v_j\right\} - \sigma_{u_k v} \sigma_{u_l v} . \\
 &= \frac{1}{n} (\sigma_{u_k u_l} \sigma_v^2 + 2\sigma_{u_k v} \sigma_{u_l v}) + \frac{n(n-1)}{n^2} \sigma_{u_k v} \sigma_{u_l v} \\
 &\quad - \sigma_{u_k v} \sigma_{u_l v} . \\
 &= \frac{1}{n} (\sigma_{u_k u_l} \sigma_v^2 + \sigma_{u_k v} \sigma_{u_l v}) .
 \end{aligned}$$

Hence, we obtain the required result. //

We now state and prove the main theorem of this section.

Theorem 4.C.1

Given assumptions i) - vi), for the cluster model with errors-in-variables, the estimator

$$\tilde{\beta}_1 = (F'F - n^* \Sigma_{uu})^{-1} F'G$$

is such that the limiting distribution of $n^{*1/2}(\tilde{\beta}_1 - \beta_1)$ is normal with mean zero and covariance matrix,

$$\bar{M}_{xx}^{*-1} \sigma_{\delta}^2 + \bar{M}_{xx}^{*-1} [\sigma_{\delta}^2 \mathbb{I}_{uu} + (1-c_*) \mathbb{I}_{uv} \mathbb{I}_{vu}] \bar{M}_{xx}^{*-1}$$

where

$$c_* = \lim_{k \rightarrow \infty} \sum_{i=1}^k \frac{n_i \sigma_{\delta}^2 \sigma_w^2}{(\sigma_{\delta}^2 + n_i \sigma_w^2)^2}$$

is assumed to exist.

Proof:

Define

$$M_{ff} = \frac{f' f}{n_*},$$

$$M_{fg} = \frac{f' g}{n_*},$$

$$a = \frac{1}{n_*} (f' \eta + \eta' f + \eta' \eta) - \mathbb{I}_{uu},$$

$$b = \frac{1}{n_*} (\eta' g + f' \zeta + \eta' \zeta).$$

Then

$$\tilde{\beta}_1 = (M_{ff} + a)^{-1} (M_{fg} + b).$$

From equations (4.C.15) through (4.C.18), we obtain

$$\nu = \zeta - \eta \beta_1 .$$

Now

$$\tilde{\beta}_1 - \beta_1 = M_{ff}^{-1} (b - a \beta_1) + O_p(n^{*-1}) \quad (4.C.20)$$

where

$$b - a \beta_1 = \frac{1}{n^*} F' \nu - \zeta_{uv} . \quad (4.C.21)$$

Substituting (4.C.21) into (4.C.20),

$$\tilde{\beta}_1 - \beta_1 = M_{ff}^{-1} \left(\frac{1}{n^*} F' \nu - \zeta_{uu} \beta_1 \right) + O_p(n^{*-1})$$

and hence,

$$\begin{aligned} (\tilde{\beta}_1 - \beta_1)(\tilde{\beta}_1 - \beta_1)' &= M_{ff}^{-1} \left(\frac{1}{n^*} F' \nu - \zeta_{uu} \beta_1 \right) \\ &\times \left(\frac{1}{n^*} F' \nu - \zeta_{uu} \beta_1 \right)' M_{ff}^{-1} + O_p(n^{*\frac{3}{2}}) . \end{aligned} \quad (4.C.22)$$

We take the expectation of

$$\left(\frac{1}{n^*} F' \nu - \zeta_{uu} \beta_1 \right) \left(\frac{1}{n^*} F' \nu - \zeta_{uu} \beta_1 \right)' :$$

$$\begin{aligned}
& E\left\{\left(\frac{1}{n^*} F'_{\mathcal{V}} - \hat{\mathcal{F}}_{uu} \beta_1\right) \left(\frac{1}{n^*} F'_{\mathcal{V}} - \hat{\mathcal{F}}_{uu} \beta_1\right)'\right\} \\
&= E\left\{\left[\frac{1}{n^*} (f'_{\mathcal{V}} + \eta'_{\mathcal{V}}) - \hat{\mathcal{F}}_{uv}\right] \left[\frac{1}{n^*} (f'_{\mathcal{V}} + \eta'_{\mathcal{V}}) - \hat{\mathcal{F}}_{uv}\right]'\right\} \\
&= \frac{1}{n^*} \sigma_{\mathcal{V}}^2 M_{ff} + E\left\{\frac{\eta'_{\mathcal{V}}}{n^*} - \hat{\mathcal{F}}_{uv}\right\} \left(\frac{\mathcal{V}'\eta}{n^*} - \hat{\mathcal{F}}_{vu}\right) \}. \quad (4.C.23)
\end{aligned}$$

Now,

$$\begin{aligned}
& E\left\{\left(\frac{\eta'_{\mathcal{V}}}{n^*} - \hat{\mathcal{F}}_{uv}\right) \left(\frac{\eta'_{\mathcal{V}}}{n^*} - \hat{\mathcal{F}}_{uv}\right)'\right\} \\
&= \frac{1}{n^*} E\left\{(\eta'_{\mathcal{V}} - n^* \hat{\mathcal{F}}_{uv})(\mathcal{V}'\eta - n^* \hat{\mathcal{F}}_{vu})\right\} \}. \quad (4.C.24)
\end{aligned}$$

Furthermore, we may decompose $\eta'_{\mathcal{V}} - n^* \hat{\mathcal{F}}_{uv}$ into the sum of k independent random variables:

$$\eta'_{\mathcal{V}} - n^* \hat{\mathcal{F}}_{uv} = \sum_{i=1}^k \left\{ \eta'_i \mathcal{V}_i - \left[(n_i - 1) + \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \right] \hat{\mathcal{F}}_{uv} \right\} \}. \quad (4.C.25)$$

Next, we partition η_i and \mathcal{V}_i into

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} \text{ and } \begin{pmatrix} \mathcal{V}_{i1} \\ \mathcal{V}_{i2} \end{pmatrix}, \quad i=1, 2, \dots, k;$$

respectively, where η_{i1} and ν_{i1} are $1 \times p$ and 1×1 vectors respectively, and, η_{i2} and ν_{i2} are $(n_i-1) \times p$ and $(n_i-1) \times 1$ matrices respectively.

Then the terms on the right-hand side of equation (4.C.25) may further be decomposed into

$$\begin{aligned} \sum_{i=1}^k [\eta_{i1}' \nu_{i1} - \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \hat{\epsilon}_{uv}] \\ + \sum_{i=1}^k [\eta_{i2}' \nu_{i2} - (n_i - 1) \hat{\epsilon}_{uv}] . \end{aligned} \quad (4.C.26)$$

Define

$$\eta_2' \nu_2 = \sum_{i=1}^k \eta_{i2}' \nu_{i2} . \quad (4.C.27)$$

Substituting definition (4.C.27) into equation (4.C.26),

$$\eta' \nu - n^* \hat{\epsilon}_{uv} = \sum_{i=1}^k (\eta_{i1}' \nu_{i1} - \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2}) + [\eta_2' \nu_2 - (n-k) \hat{\epsilon}_{uv}] . \quad (4.C.28)$$

Multiplying equation (4.C.24) by n^{*2} and substituting equation (4.C.28) into it, we obtain

$$\begin{aligned}
& \sum_{i=1}^k E(\eta'_{i1} \nu_{i1} - \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \sharp_{uv}) (\nu'_{i1} \eta_{i1} - \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \sharp_{vu}) \\
& + E[\eta'_2 \nu_2 - (n-k) \sharp_{uv}] [\nu'_2 \eta_2 - (n-k) \sharp_{vu}] . \quad (4.C.29)
\end{aligned}$$

Now, $\begin{pmatrix} \eta_{i1} \\ \nu_{i1} \end{pmatrix}$ satisfies the conditions of Lemma 4.D.2 and hence,

$$\begin{aligned}
& E(\eta'_{i1} \nu_{i1} - \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \sharp_{uv}) (\nu'_{i1} \eta_{i1} - \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \sharp_{vu}) \\
& = \sigma_{\nu_1}^2 \sharp_{\eta_1 \eta_1} + \sharp_{\eta_1 \nu_1} \sharp_{\nu_1 \eta_1} , \quad (4.C.30)
\end{aligned}$$

where

$$\begin{aligned}
\sharp_{\eta_1 \nu_1} & = E(\eta'_{i1} \nu_{i1}) \\
& = \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \sharp_{uv} ,
\end{aligned}$$

$$\sigma_{\nu_1}^2 = \sigma_{\delta}^2 \quad \text{by Lemma 4.C.1.}$$

We show that

$$E(\eta'_{i1} \nu_{i1}) = \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i \sigma_w^2} \sharp_{uv} .$$

Denote the m^{th} element of η_{i1} by η_{ilm} . Hence

$$\begin{aligned}\sigma_{\eta_{ilm} \nu_{il}} &= E(\eta_{ilm} \nu_{il}) \\ &= \frac{\sigma_{\delta}^2}{n_i(\sigma_{\delta}^2 + n_i \sigma_w^2)} E\left[\left(\sum_{j=1}^{n_i} u_{ijm}\right)\left(\sum_{l=1}^{n_i} \nu_{il}\right)\right] \\ &= \frac{\sigma_{\delta}^2}{n_i(\sigma_{\delta}^2 + n_i \sigma_w^2)} E\left[\left(\sum_{j=1}^{n_i} u_{ijm}\right)(n_i w_i + \sum_{l=1}^{n_i} \delta_{il})\right].\end{aligned}$$

Recall from equation (4.C.6) that,

$$\delta_{il} = s_{il} - u_{il} \beta_1,$$

hence

$$\begin{aligned}\sigma_{\eta_{ilm} \nu_{il}} &= - \frac{\sigma_{\delta}^2}{n_i(\sigma_{\delta}^2 + n_i \sigma_w^2)} E\left(\sum_{l=1}^{n_i} u_{il} \beta_1\right)\left(\sum_{j=1}^{n_i} u_{ijm}\right) \\ &= - \frac{\sigma_{\delta}^2}{n_i(\sigma_{\delta}^2 + n_i \sigma_w^2)} E\left(\sum_{l=1}^{n_i} \sum_{k=1}^p u_{ilk} \beta_{lk}\right)\left(\sum_{j=1}^{n_i} u_{ijm}\right) \\ &= - \frac{\sigma_{\delta}^2}{n_i(\sigma_{\delta}^2 + n_i \sigma_w^2)} E\left(\sum_{j=1}^{n_i} \sum_{k=1}^p u_{ijk} u_{ijm} \beta_{lk}\right)\end{aligned}$$

$$\begin{aligned}
&= - \frac{\sigma_{\delta}^2}{n_i(\sigma_{\delta}^2 + n_i\sigma_w^2)} \sum_{j=1}^{n_i} \left(\sum_{k=1}^p \sigma_{u_k u_m} \beta_{lk} \right) \\
&= \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i\sigma_w^2} \sigma_{u_m v} .
\end{aligned}$$

Therefore equation (4.C.30) may be written as

$$\sigma_{\delta}^2 \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i\sigma_w^2} \ddagger_{uu} + \left(\frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + n_i\sigma_w^2} \right)^2 \ddagger_{uv} \ddagger_{vu} . \quad (4.C.31)$$

We next show that $E(\eta_{i2}' \nu_{i2}) = (n-k) \ddagger_{uv}$. $\eta_{i2}' \nu_{i2}$ can be decomposed into

$$\sum_{i=1}^k \eta_{i2}' \nu_{i2} .$$

Denote the m^{th} column of η_{i2} by η_{i2m} where the elements of η_{i2m} are $(\eta_{i21m}, \eta_{i22m}, \dots, \eta_{i, 2, n_i-1, m})'$. The m^{th} element of $\eta_{i2}' \nu_{i2}$ is

$$\sum_{j=1}^{n_i-1} \eta_{i2jm} \nu_{i2m} .$$

Hence

$$E(\eta_{i2jm} \nu_{i2j}) = E\left[\frac{1}{j(j-1)} \left\{ \sum_{k=1}^{j-1} u_{ikm} - (j-1)u_{ijm} \right\} \left\{ \sum_{k=1}^{j-1} v_{ik} - (j-1)v_{ij} \right\} \right]$$

$$\begin{aligned}
&= \frac{1}{j(j-1)} E \left[\sum_{k=1}^{j-1} u_{ikm} - (j-1)u_{ijm} \right] \left[\sum_{k=1}^{j-1} (w_1 + \delta_{ik}) - (j-1)(w_1 + \delta_{ij}) \right] \\
&= \frac{-1}{j(j-1)} E \left[\sum_{k=1}^{j-1} u_{ikm} - (j-1)u_{ijm} \right] \left[\sum_{k=1}^{j-1} u_{ik} \beta_1 - (j-1)u_{ij} \beta_1 \right] \\
&= \frac{-1}{j(j-1)} E \left[\sum_{k=1}^{j-1} \sum_{l=1}^p u_{ikm} u_{ikl} \beta_{1l} + (j-1)^2 \sum_{l=1}^p u_{ijm} u_{ijl} \beta_{1l} \right] \\
&= - \sum_{l=1}^p \sigma_{u_m u_l} \beta_{1l} \\
&= \sigma_{u_m v} .
\end{aligned}$$

Hence, $E(\eta_2' \nu_2) = (n-k) \zeta_{uv}$. The rows of η_2 are independently and identically distributed. The elements of ν_2 are independently and identically distributed with variance σ_δ^2 . Therefore, using Lemma 4.C.2

$$\begin{aligned}
&E[\eta_2' \nu_2 - (n-k) \zeta_{uv}] [\nu_2' \eta_2 - (n-k) \zeta_{vu}] \\
&= (n-k) [\sigma_{\nu_2}^2 \zeta_{\eta_2 \eta_2} + \zeta_{\eta_2 \nu_2} \zeta_{\nu_2 \eta_2}] ,
\end{aligned}$$

where

$$\zeta_{\eta_2 \nu_2} = \zeta'_{\nu_2 \eta_2} = E\left(\frac{\eta_2' \nu_2}{n-k}\right) = \zeta_{uu}$$

and $\sigma_{\nu_2}^2 = \sigma_\delta^2$. Therefore

$$\begin{aligned} E[\eta_2' \nu_2 - (n-k) \mathbf{z}_{uv}] [\nu_2' \eta_2 - (n-k) \mathbf{z}_{vu}] \\ = (n-k) [\sigma_\delta^2 \mathbf{z}_{uu} + \mathbf{z}_{uv} \mathbf{z}_{vu}] . \end{aligned} \quad (4.C.32)$$

Using equations (4.C.31) and (4.C.32), we may write equation (4.C.29) as,

$$\begin{aligned} \sum_{i=1}^k \left[\sigma_\delta^2 \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \mathbf{z}_{uu} + \left(\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right) \mathbf{z}_{uv} \mathbf{z}_{vu} \right] \\ + (n-k) [\sigma_\delta^2 \mathbf{z}_{uu} + \mathbf{z}_{uv} \mathbf{z}_{vu}] . \\ = \left[(n-k) + \sum_{i=1}^k \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right] \sigma_\delta^2 \mathbf{z}_{uu} \\ + \left[(n-k) + \sum_{i=1}^k \left(\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right)^2 \right] \mathbf{z}_{uv} \mathbf{z}_{vu} , \\ = n^* [\sigma_\delta^2 \mathbf{z}_{uu} + \mathbf{z}_{uv} \mathbf{z}_{vu}] + \sum_{i=1}^k \left(\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right) \\ \times \left(\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} - 1 \right) \mathbf{z}_{uv} \mathbf{z}_{vu} \end{aligned}$$

$$= n^* [\sigma_\delta^2 \mathbb{I}_{uu} + \mathbb{I}_{uv} \mathbb{I}_{vu}] - \sum_{i=1}^k \frac{n_i \sigma_\delta^2 \sigma_w^2}{(\sigma_\delta^2 + n_i \sigma_w^2)^2} \mathbb{I}_{uv} \mathbb{I}_{vu}.$$

Equation (4.C.24) may therefore be written as

$$\begin{aligned} & \frac{1}{n^*} M_{ff} \sigma_\delta^2 + \frac{1}{n^*} (\sigma_\delta^2 \mathbb{I}_{uu} + \mathbb{I}_{uv} \mathbb{I}_{vu}) \\ & - \frac{1}{n^*} \sum_{i=1}^k \frac{n_i \sigma_\delta^2 \sigma_w^2}{(\sigma_\delta^2 + n_i \sigma_w^2)^2} \mathbb{I}_{uv} \mathbb{I}_{vu}. \end{aligned}$$

To establish the asymptotic normality of $n^{*\frac{1}{2}} (\tilde{\beta}_1 - \beta_1)$, we consider the linear combination

$$n^{*\frac{1}{2}} \lambda' M_{ff} (\tilde{\beta}_1 - \beta_1)$$

where λ is an arbitrary vector.

$$n^{*\frac{1}{2}} \lambda' M_{ff} (\tilde{\beta}_1 - \beta_1)$$

$$= n^{*\frac{1}{2}} \lambda' (b_1 - a\beta_1) + o_p(n^{*-\frac{1}{2}})$$

$$= n^{*\frac{1}{2}} \sum_{i=1}^k \sum_{j=1}^{n_i} \sum_{t=1}^p \lambda_t (f_{ijt} v_{ij} + \eta_{ijt} v_{ij} - \sigma_{u_t v}) + o_p(n^{*-\frac{1}{2}})$$

$$= S_n^* + O_p(n^{*-1/2}) .$$

The random variables $\sum_{t=1}^P \lambda_t (f_{ijt} \nu_{ij} + \eta_{ijt} \nu_{ij} - \sigma_{u_t v})$ are independent with mean zero and bounded third moments. If we denote the third moment by γ_{ij} and the second moment by θ_{ij} , then it is clear that

$$\lim_{k \rightarrow \infty} \frac{\left(\sum_{i=1}^k \sum_{j=1}^{n_i} \gamma_{ij} \right)^{1/3}}{\left(\sum_{i=1}^k \sum_{j=1}^{n_i} \theta_{ij} \right)^{1/2}} = 0 ,$$

and by Liapounov's Central Limit Theorem, S_n^* converges in distribution to a normal random variable.

Since the γ_t are arbitrary, it follows that $n^{*1/2} (\tilde{\beta}_1 - \beta_1)$ converges in distribution to a p -dimensional normal random variable. //

5. An estimator for the variance of the regression coefficients given nonnormal error structure

We now consider the case where $\epsilon_{it} = (d_{it}; u_{it})$ is not distributed as a multivariate normal. Recall that our estimator of β_1 is

$$\tilde{\beta}_1 = \left(\frac{F'F}{n} - \mathbb{I}_{uu} \right)^{-1} \frac{F'E}{n} . \quad (4.C.33)$$

$$= \left(\frac{F'F}{n} - \mathbb{I}_{uu} \right)^{-1} \left(\frac{F'F}{n} \beta_1 + \frac{F'v}{n} \right)$$

$$= \left(\frac{F'F}{n} - \mathbb{I}_{uu} \right)^{-1} \left[\left(\frac{F'F}{n} - \mathbb{I}_{uu} \right) \beta_1 + \left(\frac{F'v}{n} + \mathbb{I}_{uu} \beta_1 \right) \right] .$$

Recall that $\mathbb{I}_{uv} = -\mathbb{I}_{uu} \beta_1$. (4.C.34)

Substituting (4.C.34) into (4.C.33),

$$\tilde{\beta}_1 - \beta_1 = \left(\frac{F'F}{n} - \mathbb{I}_{uu} \right)^{-1} \left(\frac{F'v}{n} - \mathbb{I}_{uv} \right) . \quad (4.C.35)$$

We may rewrite $\frac{F'v}{n} - \mathbb{I}_{uv}$ as

$$\frac{1}{n} \sum_{i=1}^k F_i' v_i - \mathbb{I}_{uv} , \quad (4.C.36)$$

where F_i is an $(n_i \times p)$ matrix and v_i is an $(n_i \times 1)$ vector. We utilize model (4.C.1) and assume

- vii) w_i , $s_{i't}$ and $u_{i''t}$, are independent for all i, i', i'' ,
 t and t' .
- viii) The $4+\delta$, $\delta>0$, moments of d_{it} and u_{it} , are finite and
 uniformly bounded for all $i=1, 2, \dots, k$ and $t=1, 2, \dots, n_i$.
- ix) The fourth moments of x_{it} are finite and uniformly
 bounded for all $i=1, 2, \dots, k$, and $t=1, 2, \dots, n_i$.

Define

$$b_k = \frac{1}{n^*} \sum_{i=1}^k \left[(n_i - 1) + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right]^2. \quad (4.C.37)$$

Theorem 4.C.2

Given assumptions i), ii), iii), v), vi), vii), viii), and ix),

$$n^{*\frac{1}{2}} (\tilde{\beta}_1 - \beta_1) \xrightarrow{L} N(0, \bar{M}_{xx}^{*-1} \lim_{k \rightarrow \infty} H_k \bar{M}_{xx}^{*-1})$$

and

$$\left(\frac{F'F}{n^*} - \mathbb{E}_{uu} \right)^{-1} \tilde{H}_k \left(\frac{F'F}{n^*} - \mathbb{E}_{uu} \right)^{-1}$$

is a consistent estimator for the variance of β_1 , where

$$\tilde{\beta}_1 = \left(\frac{F'F}{n^*} - \mathbb{E}_{uu} \right)^{-1} \frac{F'G}{n^*}$$

$$H_k = n^* E \left[\left(\frac{F'V}{n^*} - \hat{\Phi}_{uv} \right) \left(\frac{F'V}{n^*} - \hat{\Phi}_{uv} \right)' \right]$$

$$\tilde{H}_k = \frac{1}{n^*} \sum_{i=1}^k \tilde{d}_{i..}' \tilde{d}_{i..} - b_k \tilde{\bar{d}}_{...}' \tilde{\bar{d}}_{...}$$

$$\tilde{d}_{i..} = \sum_{j=1}^{n_i} \tilde{d}_{ij.}$$

$$\tilde{\bar{d}}_{...} = \frac{1}{n^*} \sum_{i=1}^k \tilde{d}_{i..}$$

$$\tilde{d}_{ij.} = F'_{ij.} \tilde{V}_{ij}$$

$$\tilde{V}_{ij} = E_{ij} - F_{ij.} \tilde{\beta}_1$$

$$V_{ij} = E_{ij} - F_{ij.} \beta_1,$$

$F_{ij.}$ is the j^{th} row of the $n_i \times p$ matrix F_i , and E_{ij} is the j^{th} element of the $n_i \times 1$ vector E_i , $i=1, 2, \dots, k$; $j=1, 2, \dots, n_i$.

Proof:

From equation (4.C.35)

$$(\tilde{\beta}_1 - \beta_1) = \left(\frac{F'F}{n^*} - \hat{\Phi}_{uu} \right)^{-1} \left(\frac{F'V}{n^*} - \hat{\Phi}_{uv} \right)$$

$$= (M_{ff} + a)^{-1} \left(\frac{F'V}{n^*} - \hat{\Phi}_{uv} \right) \quad (4.C.38)$$

where

$$M_{ff} = \frac{f' f}{n^*}$$

$$a = \frac{1}{n^*} (f' \eta + \eta' f + \eta' \eta) - \zeta_{uu}.$$

By assumptions i), ii), and viii), $\text{var}(a) = O(n^{*-1})$ and

$$E\left(\frac{F' \mathcal{V}}{n^*} - \zeta_{uv}\right)\left(\frac{F' \mathcal{V}}{n^*} - \zeta_{uv}\right)' = O(n^{*-1}).$$

Hence, by Corollary 2.A.2, $a = O_p(n^{*-1/2})$ and

$$\frac{F' \mathcal{V}}{n^*} - \zeta_{uv} = O_p(n^{*-1/2}).$$

We may write equation (4.C.35) as

$$\tilde{\beta}_1 - \beta_1 = M_{ff}^{-1} \left(\frac{F' \mathcal{V}}{n^*} - \zeta_{uv} \right) + O_p(n^{*-1/2}). \quad (4.C.39)$$

From Theorem 2.B.1, it follows that the limiting distribution of

$n^{*1/2} (\tilde{\beta}_1 - \beta_1)$ is the same as the limiting distribution of

$$n^{*1/2} M_{ff}^{-1} \left(\frac{F' \mathcal{V}}{n^*} - \zeta_{uv} \right).$$

Let ρ be an arbitrary nonzero $(p \times 1)$ vector and consider

$$n^{*\frac{1}{2}} \rho' \left(\frac{F' \mathcal{V}}{n^*} - \mathcal{Z}_{uv} \right) .$$

Now,

$$\begin{aligned} \frac{F' \mathcal{V}}{n^*} - \mathcal{Z}_{uv} &= \frac{1}{n^*} (F' \mathcal{V} - n^* \mathcal{Z}_{uv}) \\ &= \frac{1}{n^*} \left(\sum_{i=1}^k F'_i \mathcal{V}_i - \sum_{i=1}^k \left[n_i - 1 + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right] \mathcal{Z}_{uv} \right) \\ &= \frac{1}{n^*} \sum_{i=1}^k (F'_i \mathcal{V}_i - [n_i - 1 + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2}] \mathcal{Z}_{uv}) . \end{aligned} \quad (4.C.40)$$

Define

$$\theta_i = F'_i \mathcal{V}_i - (n_i - 1 + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2}) \mathcal{Z}_{uv} .$$

Then we may write equation (4.C.40) as the sum of k independent random variables:

$$\frac{1}{n^*} \sum_{i=1}^k \theta_i .$$

It follows that

$$E \left[\frac{1}{n^*} \sum_{i=1}^k \theta_i \right] = 0 .$$

Note that $E[(\rho' \theta_i)^2]$ and $E[|\rho' \theta_i|^{2+1/\delta}]$, $\delta > 0$, are bounded by the boundedness of the x 's and the finite $4+\delta$ moments. Hence

$$\lim_{k \rightarrow \infty} \frac{\sum_{i=1}^k E[|\rho' \theta_i|^{2+1/\delta}]}{\left\{ \sum_{i=1}^k E[(\rho' \theta_i)^2] \right\}^{1+1/4\delta}} = 0.$$

By the Liapounov central limit theorem, Theorem 2.C.2, this gives

$$\frac{\sum_{i=1}^k \rho' \theta_i}{\left\{ \rho' E \left[\sum_{i=1}^k (\theta_i \theta_i') \right] \rho \right\}^{1/2}} \xrightarrow{L} N(0,1)$$

or

$$\frac{n^{*-1/2} \sum_{i=1}^k \rho' \theta_i}{\left\{ \rho' E \left[\frac{1}{n^*} \sum_{i=1}^k (\theta_i \theta_i') \right] \rho \right\}^{1/2}} \xrightarrow{L} N(0,1).$$

By a result in Rao [1965, p. 102],

$$n^{*-1/2} \sum_{i=1}^k \rho' \theta_i \xrightarrow{L} N(0, \rho' \lim_{k \rightarrow \infty} E \left[\frac{1}{n^*} \sum_{i=1}^k \theta_i \theta_i' \right] \rho).$$

We proceed to prove that

$$E \left[\frac{1}{n^*} \sum_{i=1}^k \theta_i \theta_i' \right]$$

is H_k . We have that

$$\begin{aligned}
 H_k &= n^* E \left\{ \left(\frac{F'_i \nu_i}{n^*} - \bar{x}_{uv} \right) \left(\frac{F'_i \nu_i}{n^*} - \bar{x}_{uv} \right)' \right\} \\
 &= n^* E \left\{ \left(\sum_{i=1}^k \frac{F'_i \nu_i}{n^*} - \bar{x}_{uv} \right) \left(\sum_{i=1}^k \frac{F'_i \nu_i}{n^*} - \bar{x}_{uv} \right)' \right\} \\
 &= n^{*-1} E \left\{ \left[\sum_{i=1}^k (F'_i \nu_i - \left(\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} + n_i - 1 \right) \bar{x}_{uv}) \right] \right. \\
 &\quad \left. \left[\sum_{i=1}^k (F'_i \nu_i - \left(\frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} + n_i - 1 \right) \bar{x}_{uv}) \right]' \right\} \\
 &= n^{*-1} E \left\{ \sum_{i=1}^k \left[F'_i \nu_i - \left(n_i - 1 + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right) \bar{x}_{uv} \right] \right. \\
 &\quad \left. \left[F'_i \nu_i - \left(n_i - 1 + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right) \bar{x}_{uv} \right]' \right\} \\
 &= n^{*-1} E \left\{ \sum_{i=1}^k \theta_i \theta_i' \right\}.
 \end{aligned}$$

Hence, by the multivariate central limit theorem, we now have

$$n^{*-1/2} \sum_{i=1}^k \theta_i \xrightarrow{L} N(0, \lim_{k \rightarrow \infty} H_k).$$

Finally, since $\lim_{n \rightarrow \infty} M_{ff}^* = \bar{M}_{xx}^*$ where \bar{M}_{xx}^* is nonsingular, we obtain

$$M_{ff}^{-1} n^{*-1/2} \sum_{i=1}^k \theta_i \xrightarrow{L} N(0, \bar{M}_{xx}^{*-1} \lim_{k \rightarrow \infty} H_k \bar{M}_{xx}^{*-1}) .$$

We now prove that \tilde{H}_k is a consistent estimator of H_k . Recall that

$$\tilde{H}_k = \frac{1}{n^*} \sum_{i=1}^k \tilde{d}_{i..}' \tilde{d}_{i..} - b_k \bar{d}_{...}' \bar{d}_{...} , \quad (4.C.41)$$

and

$$\tilde{\nu}_i = \nu_i - F_i(\tilde{\beta}_1 - \beta_1) .$$

The $(r,s)^{th}$ element of $\tilde{d}_{i..}' \tilde{d}_{i..}$ is

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^k \left[\sum_{j=1}^{n_i} F_{ijr} \tilde{\nu}_{ij} \right] \left(\sum_{l=1}^{n_i} F_{il s} \tilde{\nu}_{il} \right) \\ &= \frac{1}{n^*} \left[\sum_{j=1}^{n_i} F_{ijr} [\nu_{ij} - F_{ij} \cdot (\tilde{\beta}_1 - \beta_1)] \right. \\ & \quad \left. \left[\sum_{l=1}^{n_i} F_{il s} [\nu_{il} - F_{il} \cdot (\tilde{\beta}_1 - \beta_1)] \right] \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^*} \sum_{i=1}^k \left(\sum_{j=1}^{n_i} F_{ijr} \nu_{ij} \right) \left(\sum_{l=1}^{n_i} F_{ils} \nu_{il} \right) \\
&\quad - \frac{1}{n^*} \sum_{i=1}^k \left[\sum_{j=1}^{n_i} F_{ijr} \nu_{ij} \right] \left(\sum_{j=1}^{n_i} F_{ils} F_{il.} \right) (\tilde{\beta}_1 - \beta_1) \\
&\quad - \frac{1}{n^*} \sum_{i=1}^k \left[\left(\sum_{l=1}^{n_i} F_{ils} \nu_{il} \right) \left(\sum_{j=1}^{n_i} F_{ijr} F_{ij.} \right) \right] (\tilde{\beta}_1 - \beta_1) \\
&\quad + (\tilde{\beta}_1 - \beta_1) \left[\frac{1}{n^*} \sum_{i=1}^k \left(\sum_{j=1}^{n_i} F_{ijr} F_{ij.} \right) \left(\sum_{l=1}^{n_i} F_{il.} F_{ils} \right) \right] (\tilde{\beta}_1 - \beta_1) .
\end{aligned} \tag{4.C.42}$$

Since $(\tilde{\beta}_1 - \beta_1) = O_p(n^{*-1/2})$ and using assumption ix), we obtain that

$$\tilde{d}_{i..}' \tilde{d}_{i..} = \frac{1}{n^*} \sum_{i=1}^k (F_i' \nu_i) (F_i' \nu_i) + O_p(n^{*-1/2}) . \tag{4.C.43}$$

Also

$$\tilde{d}_{...} = \frac{1}{n^*} \sum_{i=1}^k \{ F_i' \nu_i + F_i' F_i (\tilde{\beta}_1 - \beta_1) \} ,$$

and

$$\begin{aligned}
\bar{d}' \dots \bar{d}' &= \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right)' \\
&+ \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' F_i (\tilde{\beta}_1 - \beta_1) \right)' \\
&+ \left(\frac{1}{n^*} \sum_{i=1}^k F_i' F_i (\tilde{\beta}_1 - \beta_1) \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right)' \\
&+ \left(\frac{1}{n^*} \sum_{i=1}^k F_i' F_i (\tilde{\beta}_1 - \beta_1) \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' F_i (\tilde{\beta}_1 - \beta_1) \right)' .
\end{aligned}$$

Since

$$\frac{1}{n^*} \sum_{i=1}^k F_i' F_i = O(1)$$

$$\text{and } (\tilde{\beta}_1 - \beta_1) = O_p(n^{*-1/2}),$$

$$\bar{d}' \dots \bar{d}' = \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right)' + O_p(n^{*-1/2}) . \quad (4.C.44)$$

In view of (4.C.43) and (4.C.44) we write equation (4.C.41) as

$$\begin{aligned}
\tilde{H}_k &= \frac{1}{n^*} \sum_{i=1}^k (F_i' \nu_i) (F_i' \nu_i)' \\
&- b_k \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right)' + O_p(n^{*-1/2})
\end{aligned}$$

$$= \hat{H}_k + O_p(n^{*-1/2}), \quad (4.C.45)$$

where

$$\hat{H}_k = \frac{1}{n^*} \sum_{i=1}^k (F_i' \nu_i)(F_i' \nu_i)' - b_k \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right) \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right)' . \quad (4.C.46)$$

Now, $\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i$ is the sum of k independent random variables

with each of its components (p of them) having second moments.

Since

$$E \left(\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i \right) = \Sigma_{uv} ,$$

we get that

$$\frac{1}{n^*} \sum_{i=1}^k F_i' \nu_i = \Sigma_{uv} + O_p(n^{*-1/2}) . \quad (4.C.47)$$

$\frac{1}{n^*} \sum_{i=1}^k (F_i' \nu_i)(F_i' \nu_i)'$ is the sum of k independent random variables

with uniformly bounded $1+\frac{1}{2}\delta$ moments. Hence by Theorem 3.D.1,

$$\frac{1}{n^*} \left[\sum_{i=1}^k \{ (F_i' \nu_i)(F_i' \nu_i)' - E(F_i' \nu_i)(F_i' \nu_i)' \} \right] \xrightarrow{P} 0 . \quad (4.C.48)$$

Also

$$\begin{aligned}
 H_k &= n^{*-1} \left[E \sum_{i=1}^k (F_i' \nu_i) (F_i' \nu_i)' \right. \\
 &\quad \left. - \sum_{i=1}^k \left(n_i - 1 + \frac{\sigma_\delta^2}{\sigma_\delta^2 + n_i \sigma_w^2} \right)^2 \zeta_{uv} \zeta_{vu} \right] . \\
 &= n^{*-1} E \sum_{i=1}^k (F_i' \nu_i) (F_i' \nu_i)' - b_k \zeta_{uv} \zeta_{vu} . \quad (4.C.49)
 \end{aligned}$$

Combining equations (4.C.47) and (4.C.46), we obtain

$$\tilde{H}_k = \frac{1}{n^*} \sum_{i=1}^k (F_i' \nu_i) (F_i' \nu_i)' - b_k \zeta_{uv} \zeta_{vu} + O_p(n^{*-1/2}) . \quad (4.C.50)$$

Combining equations (4.C.49) and (4.C.50), we obtain

$$\tilde{H}_k - H_k \xrightarrow{P} 0 \quad . \quad //$$

D. Estimation when the Error Variance is a
Multiple of Total Variance

1. Introduction

Cochran [1968, 1970] and Wiley [1973] have dealt with the distortions that may be introduced into standard least-squares estimating procedures by measurement error in the independent variables of a regression equation. Johnson [1963], Walker and Lev [1953] have remarked that the regression coefficient is attenuated (reduced in absolute value) when compared with the coefficient computed in absence of measurement error.

Blalock et al. [1969] and Curtiss and Jackson [1962] made use of multiple indicators to estimate measurement error. These estimates of measurement error were used for attenuation correction as a modification of the least-squares estimation. One problem, as pointed out by Bohrnstedt and Carter [1971, p. 142], with correction for attenuation has been the lack of sampling theory for the estimates.

In the traditional errors-in-variables literature the error variance σ_u^2 is assumed known. In the correction for attenuation approach the ratio of the error variance to the total variance of the independent variable,

$\frac{\sigma_u^2}{\sigma_x^2}$, is assumed known. We develop some theory for this case.

2. The model and assumptions

Consider the following errors-in-variables regression model

$$y = x \beta ,$$

where

$$X = x + u , \quad (4.D.1)$$

$$Y = y + \epsilon ,$$

and (X,Y) can be observed. The properties of our estimator of β rest on the following assumptions:

- i) x is an $(n \times p)$ matrix of random variables, whose rows $x_{t.}$, $t=1, 2, \dots, n$, are independently and identically distributed as a multivariate normal with mean zero and covariance matrix Σ_{xx} , Σ_{xx} nonsingular.
- ii) $(\epsilon; u)$ is an $n \times (p+1)$ matrix of random variables whose rows $(\epsilon_t; u_{t.})$, $t=1, 2, \dots, n$, are independently and identically distributed as a multivariate normal with mean 0 and positive definite covariance matrix,

$$\Sigma = \begin{pmatrix} \sigma_{\epsilon}^2 & 0 \\ 0 & D_{uu} \end{pmatrix}$$

where D_{uu} is a $p \times p$ diagonal matrix.

- iii) The normal vector $(y_t; x_t.)$ is independent of the error vector $(\epsilon_s; u_s)$ for all t and s .
- iv) The ratio of the i^{th} diagonal element of D_{uu} to the i^{th} diagonal element of Σ_{xx} is known,

$$\lambda_i = \frac{\sigma_{u_i}^2}{\sigma_{x_i}^2}, \quad i=1, 2, \dots, p; \quad (4.D.2)$$

where each λ_i is a real nonnegative number.

2. Estimation

From equation (4.D.1), we have

$$\begin{aligned} E\left(\frac{X'X}{n}\right) &= \Sigma_{XX} \\ &= \Sigma_{xx} + D_{uu}, \end{aligned} \quad (4.D.3)$$

and

$$\begin{aligned} E\left(\frac{X'Y}{n}\right) &= \Sigma_{XY} \\ &= \Sigma_{xy} \\ &= \Sigma_{xx} \beta. \end{aligned} \quad (4.D.4)$$

Define the $1 \times (p+1)$ vector $Z_t.$ by setting

$$Z_{t.} = (Y_t; X_{t.}) , \quad t=1, 2, \dots, n .$$

Now using assumptions i) - iii), $Z_{t.}$ is distributed as a multivariate normal with mean zero and covariance matrix Σ_{ZZ} , where

$$\begin{aligned} \Sigma_{ZZ} &= \begin{pmatrix} \sigma_e^2 + \beta' \Sigma_{xx} \beta & \Sigma_{xx} \beta \\ \beta' \Sigma_{xx} & \Sigma_{xx} + D_{uu} \end{pmatrix} \\ &= \begin{pmatrix} \beta' \Sigma_{xx} \beta & \Sigma_{xx} \beta \\ \beta' \Sigma_{xx} & \Sigma_{xx} \end{pmatrix} + \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & D_{uu} \end{pmatrix} . \end{aligned}$$

Note that our errors-in-variables model fits in the class of models defined by Jöreskog [1970]. Jöreskog [1970] considered a data matrix $Z = \{Z_{tj}\}$ of n observations on $(p+1)$ response variables; $t=1, 2, \dots, n$ and $j=1, 2, \dots, p+1$, with the rows of Z , $Z_{t.}$, independently distributed with mean vector $E(Z_{t.})$ and covariance matrix Σ_{ZZ} . He assumed Σ_{ZZ} to be of the form

$$\Sigma_{ZZ} = B(M \otimes M' + \Psi^2) B' + \theta^2 ,$$

and $E(Z) = ACD$, where $A = \{a_{ij}\}_{n \times p}$ with $\text{rank}(A) = g$ and

$D = \{d_{ij}\}_{h \times (p+1)}$ with $\text{rank}(D) = h$ are both fixed matrices with

$g \leq n$ and $h \leq (p+1)$;

$$C = \{\zeta_{ij}\}_{g \times h}, B = \{\beta_{ij}\}, M = \{m_{ij}\},$$

the symmetric matrix $\Phi = \{\phi_{ij}\}$ and the diagonal matrices $\Psi = \text{diag}\{\phi_k\}$ and $\theta = \text{diag}\{\theta_i\}$ are parameter matrices. He allowed for any of the parameters in C, B, M, Φ , and θ to be known a priori and for one or more subsets of the remaining parameters to have identical but unknown values. Thus parameters were of three kinds:

- i) fixed parameters that have been assigned given values;
- ii) constrained parameters that are unknown but equal to one or more other parameters; and
- iii) free parameters that are unknown and not constrained to be equal to any other parameter.

In our case,

$$B = M = I_{p+1}, \quad \Psi = \begin{pmatrix} \sigma_e^2 & O_p \\ O_p & D_{uu} \end{pmatrix}$$

and $\theta = O_{p+1}$. I_{p+1} and O_{p+1} denote $(p+1) \times (p+1)$ unit and zero matrices respectively. Furthermore,

$$\Phi = \begin{pmatrix} \beta' \zeta_{xx} \beta & \zeta_{xx} \beta \\ \beta' \zeta_{xx} & \zeta_{xx} \end{pmatrix}$$

is a symmetric matrix. $\sigma_{x_i}^2$ and $\sigma_{u_i}^2$ are constrained according to equation (4.D.2).

We first consider the maximum likelihood estimation of the covariance matrix of the multivariate normal

$$Z_{t.} \sim N(0, \Sigma_{ZZ})$$

given the parameter space for Σ_{ZZ} is the space of all real positive definite covariance matrices.

To obtain the maximum likelihood estimator we maximize the likelihood function

$$\begin{aligned} \log L = & -\frac{1}{2} (p+1) n \log (2\pi) - \frac{1}{2} \log |\Sigma_{ZZ}| \\ & - \frac{1}{2} \sum_{t=1}^n \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} Z_{ti} \sigma^{ij} Z_{tj}, \end{aligned}$$

with respect to σ_{ij} (the elements of Σ_{ZZ}), where σ^{ij} are the elements of Σ_{ZZ}^{-1} . Maximizing $\log L$ is equivalent to maximizing

$$F = \frac{1}{2} n \log |\Sigma_{ZZ}|^{-1} - \text{trace} (\Sigma_{ZZ}^{-1} A),$$

where

$$A = \sum_{t=1}^n Z_{t.}' Z_{t.}$$

By Lemmas 3.2.2 and 3.2.3 in Anderson [1958, p. 46 and p. 47], it follows that the maximum likelihood estimator of $\hat{\Sigma}_{ZZ}$ is $\frac{A}{n}$ where A is given by

$$A = \begin{bmatrix} \sum_{t=1}^n Y_t^2, & \sum_{t=1}^n X_{t.} Y_t \\ \sum_{t=1}^n X_{t.}' Y_t, & \sum_{t=1}^n X_{t.}' X_{t.} \end{bmatrix}$$

In our case $\hat{\Sigma}_{ZZ}$ is defined in terms of the parameters of interest by,

$$\hat{\Sigma}_{ZZ} = \begin{pmatrix} \beta' \hat{\Sigma}_{xx} \beta + \sigma_e^2, & \hat{\Sigma}_{xx} \beta \\ \beta' \hat{\Sigma}_{xx}, & \hat{\Sigma}_{xx} + \Gamma \hat{\Sigma}_{xx} \end{pmatrix} \quad (4.D.5)$$

where Γ is a (p x p) diagonal matrix with diagonal elements equal to the known ratios $\lambda_1, \lambda_2, \dots, \lambda_p$. This definition means that there are certain restrictions on the form of $\hat{\Sigma}_{ZZ}$. Ignoring the restrictions, we create estimations of the parameters by equating the elements of $\hat{\Sigma}_{ZZ}$ to the corresponding elements of $\frac{A}{n}$;

$$\tilde{\beta}' \tilde{\mathbf{I}}_{xx} \tilde{\beta} + \tilde{\sigma}_e^2 = \frac{1}{n} \sum_{t=1}^n y_t^2 ,$$

$$\tilde{\beta} \tilde{\mathbf{I}}_{xx} = \frac{1}{n} \sum_{t=1}^n x_t' y_t , \quad (4.D.6)$$

$$\tilde{\mathbf{I}}_{xx} + \Gamma \tilde{\mathbf{I}}_{xx} = \frac{1}{n} \sum_{t=1}^n x_t' x_t .$$

Solving the system of equations (4.D.6), we obtain;

$$\tilde{\beta} = (\hat{\mathbf{M}}_{xx} - \Lambda \hat{\mathbf{D}}_{xx})^{-1} \hat{\mathbf{M}}_{xy} \quad (4.D.7)$$

$$\tilde{\mathbf{I}}_{xx} = (\hat{\mathbf{M}}_{xx} - \Lambda \hat{\mathbf{D}}_{xx}) \quad (4.D.8)$$

$$\tilde{\sigma}_e^2 = \frac{1}{n} \sum_{t=1}^n y_t^2 - \tilde{\beta}' \tilde{\mathbf{I}}_{xx} \tilde{\beta} , \quad (4.D.9)$$

where

$$\hat{\mathbf{D}}_{xx} = \text{mdia} \hat{\mathbf{M}}_{xx} ,$$

$\text{mdia} \hat{\mathbf{I}}_{xx}$ is the diagonal matrix whose diagonal is the diagonal of the matrix $\hat{\mathbf{I}}_{xx}$,

$$\hat{\mathbf{M}}_{xx} = \frac{1}{n} \sum_{t=1}^n x_t' x_t .$$

$$\hat{M}_{XY} = \frac{1}{n} \sum_{t=1}^n x_t' y_t$$

$$\hat{M}_{YY} = \frac{1}{n} \sum_{t=1}^n y_t^2$$

and

$$\Lambda = \begin{bmatrix} \frac{\lambda_1}{1+\lambda_1} & , & 0 & , & \dots & , & 0 \\ 0 & , & \frac{\lambda_2}{1+\lambda_2} & , & \dots & , & 0 \\ \vdots & & & & & & \vdots \\ 0 & , & 0 & , & \dots & , & \frac{\lambda_p}{1+\lambda_p} \end{bmatrix} . \quad (4.D.10)$$

The restrictions associated with equation (3.D.5), and assumptions i) and ii) are:

$$a) \quad |\tilde{\lambda}_{xx}| > 0 ,$$

$$b) \tilde{\sigma}_e^2 > 0.$$

If the estimates (4.D.7), (4.D.8) and (4.D.9) meet restrictions a) and b), then they are the maximum likelihood estimates. We shall take equations (4.D.7) - (4.D.9) as our estimators. We demonstrate that, for large n , the restriction b) is met with high probability by these estimators. Let $\hat{\theta}$ denote the column vector obtained by arranging \hat{M}_{XY} , the p columns of \hat{M}_{XX} and the scalar \hat{M}_{YY} in a single column; and let θ be the corresponding parameter vector. Since all positive moments exist for normal random variables, we have that $E\{|\hat{\theta} - \theta|^{2r}\} = O(n^{-r})$ for r a positive integer. Using Tchebychev's inequality,

$$P\{|\hat{\theta} - \theta| > e\} \leq \frac{E\{|\hat{\theta} - \theta|^{2r}\}}{e^{2r}}$$

for $e > 0$. Hence, $P\{|\hat{\theta} - \theta| > e\} = O(n^{-r})$. Now, $\tilde{\sigma}_e^2$ is a continuous differentiable function of $\hat{\theta}$ for all $\hat{\theta}$ in a bounded set containing the true θ . Hence, $P\{\tilde{\sigma}_e^2 < 0\} = O(n^{-r})$. Similar arguments hold for $|\tilde{\Sigma}_{XX}|$.

Model (4.D.1) may also be written as

$$Y = X \beta + v, \quad (4.D.11)$$

where

$$v = \epsilon - u \beta. \quad (4.D.12)$$

Using (4.D.12), the $p \times 1$ covariance vector between u and v , which we denote by Σ_{uv} is

$$\begin{aligned} \Sigma_{uv} &= \Sigma_{u\epsilon} - \Sigma_{uu} \beta \\ &= - \Sigma_{uu} \beta, \end{aligned} \quad (4.D.13)$$

since $\Sigma_{u\epsilon}$ was assumed to be zero.

We denote the elements of the covariance vector Σ_{uv} by $\sigma_{u_i v}$, $i=1, 2, \dots, p$.

4. Variance of the regression coefficients given normal error structure

Theorem 4.D.1

Given assumptions i) - v) and model (4.D.1),

$$n^{\frac{1}{2}}(\tilde{\beta} - \beta) \xrightarrow{L} N(0, \Sigma_{xx}^{-1} v \Sigma_{xx}^{-1})$$

where

$$v = (\sigma_v^2 \mathbf{I}_{XX} + \mathbf{I}_{XV} \mathbf{I}_{VX}) - A .$$

A is a (p x p) matrix with elements

$$a_{ii} = 2 \Lambda_i^2 \beta_i^2 \sigma_{X_i}^4 \quad i=1, 2, \dots, p;$$

$$a_{ij} = -2\sigma_{X_i X_j} (\Lambda_j^2 \beta_j^2 \sigma_{X_j}^2 + \Lambda_i^2 \beta_i^2 \sigma_{X_i}^2 + \Lambda_i \Lambda_j \beta_i \beta_j \sigma_{X_i X_j}) \quad i \neq j=1, 2, \dots, p .$$

$$\mathbf{I}_{XV} = -D_{uu} \beta_1 .$$

Proof:

The estimator is given by

$$\tilde{\beta} = (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} \hat{M}_{XY} . \quad (4.D.14)$$

Define $\hat{M}_{XV} = \frac{1}{n} X'v$, where v was defined in equation (4.D.12). Then

$$\hat{M}_{XY} = \hat{M}_{XX} \beta + \hat{M}_{XV} . \quad (4.D.15)$$

Using equations (4.D.14) and (4.D.15);

$$\begin{aligned} \tilde{\beta} &= (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} (\hat{M}_{XX} + \hat{M}_{XV}) \\ &= (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} [(\hat{M}_{XX} \beta - \Lambda \hat{D}_{XX} \beta) + (\Lambda \hat{D}_{XX} \beta + \hat{M}_{XV})] \end{aligned}$$

$$= \beta + (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta) . \quad (4.D.16)$$

and

$$\tilde{\beta} - \beta = (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta) . \quad (4.D.17)$$

Now,

$$\begin{aligned} E (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta) \\ &= E (\hat{M}_{XV}) + \Lambda E (\hat{D}_{XX}) \beta \\ &= E \left(\frac{1}{n} X' v \right) + \Lambda E [\text{mdia}g \left(\frac{1}{n} X' X \right)] \beta \\ &= E \left[\frac{1}{n} (x' + u') (e - u \beta) \right] + \Lambda E [\text{mdia}g \frac{1}{n} (x + u)' (x + u)] \beta \\ &= - D_{uu} \beta + \Lambda (\text{mdia}g \frac{1}{n} X' X + D_{uu}) \beta \\ &= 0 . \end{aligned}$$

We find the covariance matrix of $\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta$,

$$\begin{aligned} E (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta) (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta)' \\ &= \text{var } \hat{M}_{XV} + \text{cov} [\hat{M}_{XV}, \Lambda \hat{D}_{XX} \beta] \\ &\quad + \text{cov} [\Lambda \hat{D}_{XX} \beta, \hat{M}_{XV}] + \text{var} [\Lambda \hat{D}_{XX} \beta] . \end{aligned}$$

Let

$$\frac{X'_{i.v}}{n}, \quad \frac{X'_{j.v}}{n}$$

be the i^{th} and j^{th} elements of \hat{M}_{XV} , $i, j=1, 2, \dots, p$;

$$\Lambda_i \frac{X'_{i.X_i}}{n} \beta_i \quad \text{and} \quad \Lambda_j \frac{X'_{j.X_j}}{n} \beta_j$$

be the i^{th} and j^{th} elements of $\Lambda_{XX}^D \beta$, where $X_{.j}$ is the j^{th} column of the matrix X , $j=1, 2, \dots, p$.

Using properties of normal covariances,

$$\text{var} \left(\frac{X'_{i.v}}{n} \right) = \frac{1}{n} (\sigma_{X_i}^2 \sigma_v^2 + \sigma_{X_i.v}^2),$$

$$\text{cov} \left(\frac{X'_{i.v}}{n}, \frac{X'_{i.X_i} \beta_i \Lambda_i}{n} \right) = \frac{1}{n} \Lambda_i \beta_i (2\sigma_{X_i}^2 \sigma_{X_i.v}),$$

$$\text{var} \left(\Lambda_i \frac{X'_{i.X_i}}{n} \beta_i \right) = \frac{1}{n} \Lambda_i^2 \beta_i^2 (2\sigma_{X_i}^4),$$

$$\text{cov} \left(\frac{X'_{i.v}}{n}, \frac{X'_{j.v}}{n} \right) = \frac{1}{n} (\sigma_{X_i.X_j} \sigma_v^2 + \sigma_{X_i.v} \sigma_{X_j.v}),$$

$$\text{cov} \left(\frac{X'_{i.v}}{n}, \Lambda_j \frac{X'_{j.X_j}}{n} \beta_j \right) = \frac{1}{n} \Lambda_j \beta_j (2\sigma_{X_i.X_j} \sigma_{X_i.v}),$$

$$\text{cov} \left(\Lambda_i \frac{X'_{i.X_i}}{n} \beta_i, \frac{X'_{j.v}}{n} \right) = \frac{1}{n} \Lambda_i \beta_i (2\sigma_{X_i.X_j} \sigma_{X_j.v}),$$

$$\text{cov} \left(\Lambda_i \beta_i \frac{x_i' x_i}{n}, \Lambda_j \beta_j \frac{x_j' x_j}{n} \right) = \frac{1}{n} \Lambda_i \Lambda_j \beta_i \beta_j (2\sigma_{x_i x_j}) .$$

The i^{th} diagonal element of $\text{var} (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta)$ is:

$$\begin{aligned} \text{var} \left[\frac{1}{n} (x_i' v + \Lambda_i x_i' x_i \beta_i) \right] &= \text{var} \left(\frac{1}{n} x_i' v \right) \\ &+ 2 \text{cov} \left(\frac{1}{n} x_i' v, \frac{\Lambda_i x_i' x_i \beta_i}{n} \right) + \text{var} \left(\Lambda_i \frac{x_i' x_i}{n} \beta_i \right) \\ &= \frac{1}{n} [(\sigma_{x_i}^2 \sigma_v^2 + \sigma_{x_i v}^2) + 4 \Lambda_i \beta_i \sigma_{x_i}^2 \sigma_{x_i v} + 2 \Lambda_i^2 \beta_i^2 \sigma_{x_i}^4] . \end{aligned}$$

(4.D.18)

The $(i,j)^{\text{th}}$ off-diagonal element of $\text{var} (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta)$ is:

$$\begin{aligned} \text{cov} \left[\frac{1}{n} (x_i' v + \Lambda_i x_i' x_i \beta_i), \frac{1}{n} (x_j' v + \Lambda_j x_j' x_j \beta_j) \right] \\ = \frac{1}{n} [(\sigma_v^2 \sigma_{x_i x_j} + \sigma_{x_i v} \sigma_{x_j v}) + 2(\Lambda_j \sigma_{x_i x_j} \sigma_{x_j v} \beta_j \\ + \Lambda_i \sigma_{x_i x_j} \sigma_{x_i v} \beta_i + \Lambda_i \Lambda_j \sigma_{x_i x_j}^2 \beta_i \beta_j)] . \end{aligned}$$

(4.D.19)

Using the fact that,

$$\sigma_{u_i}^2 = \lambda_i \sigma_{x_i}^2 \tag{4.D.20}$$

and

$$\dot{\beta}_{uv} = -\dot{\beta}_{uu} \beta_i,$$

we have

$$\begin{aligned}\sigma_{u_i v} &= -\sigma_{u_i}^2 \beta_i \\ &= -\lambda_i \sigma_{x_i}^2 \beta_i, \quad i=1, 2, \dots, p.\end{aligned}\tag{4.D.21}$$

$$\begin{aligned}\sigma_{x_i v} &= E(X_i v) \\ &= \sigma_{u_i v} \\ &= -\lambda_i \sigma_{x_i}^2 \beta_i.\end{aligned}\tag{4.D.22}$$

Substituting (4.D.21) and (4.D.20) into (4.D.18), the i^{th} diagonal element of $\text{var}(\hat{M}_{Xv} + \hat{\Lambda}_{XX} \beta)$ may be written as:

$$\begin{aligned}&\frac{1}{n} \sigma_{X_i}^2 \sigma_v^2 + \left\{ \sigma_{u_i}^4 \beta_i^2 - 4 \frac{\lambda_i^2}{1+\lambda_i} \beta_i^2 (1+\lambda_i) \sigma_{x_i}^4 \right. \\ &\quad \left. + 2 \left(\frac{\lambda_i}{1+\lambda_i} \right)^2 \beta_i^2 \sigma_{x_i}^4 (1+\lambda_i)^2 \right\} \\ &= \frac{1}{n} (\sigma_{X_i}^2 \sigma_v^2 - \sigma_{u_i v}^2)\end{aligned}$$

$$= \frac{1}{n} (\sigma_{X_i}^2 \sigma_v^2 - \sigma_{X_i v}^2)$$

$$= \frac{1}{n} (\sigma_{X_i}^2 \sigma_v^2 + \sigma_{X_i v}^2) - \frac{2}{n} \sigma_{X_i v}^2 .$$

Hence the covariance matrix V of $\hat{M}_{XV} + \hat{\Delta}_{XX} \beta$ may be written

as

$$V = \frac{1}{n} (\sigma_v^2 \mathbf{I}_{XX} + \mathbf{I}_{XV} \mathbf{I}_{VX}) - \frac{A}{n}$$

where

$$\mathbf{I}_{XX} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_2 X_1} & \cdots & \sigma_{X_1 X_p} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_p X_1} & \sigma_{X_p X_2} & \cdots & \sigma_{X_p}^2 \end{bmatrix}$$

$$\mathbf{I}_{XV} = \begin{bmatrix} \sigma_{X_1 v} \\ \sigma_{X_2 v} \\ \vdots \\ \sigma_{X_p v} \end{bmatrix}$$

$$A = \{a_{ij}\}_{p \times p}$$

$$a_{ii} = 2 \sigma_{X_i v}^2$$

$$a_{ij} = -2 \sigma_{X_i X_j} (\Lambda_j \beta_j \sigma_{X_j v} + \Lambda_i \Lambda_j \beta_i \beta_j \sigma_{X_i X_j} + \Lambda_i \beta_i \sigma_{X_i v})$$

for $i \neq j = 1, 2, \dots, p$ and $i \neq j$.

Now,

$$E(\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}}) = \xi_{XX}$$

and the variance of the elements of $(\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}})$ are $O(\frac{1}{n})$.

Hence, using the Weak Law of Large Numbers,

$$(\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}}) \xrightarrow{P} \xi_{XX}.$$

We consider the linear combination,

$$n^{\frac{1}{2}} p' (\hat{M}_{Xv} + \hat{\Lambda}_{D_{XX}} \beta)$$

$$= n^{-\frac{1}{2}} \sum_{t=1}^n \sum_{i=1}^p \rho_i (x_{ti} v_t + \frac{\lambda_i}{1+\lambda_i} x_{ti}^2 \beta_i)$$

$$= S_n.$$

Now, the random variables

$$\sum_{i=1}^p \rho_i (x_{ti} v_t + x_{ti}^2 \beta_i)$$

are independently and identically distributed for normal X :
furthermore, they have zero mean. S_n converges in distribution to a normal random variable using Theorem 6.4.4, Chung [1968, p. 157].

Since the ρ_i are arbitrary, it follows that the p -dimensional vector $n^{\frac{1}{2}} (\hat{M}_{XV} + \hat{\Delta}_{XX} \beta)$ converges in distribution to a p -dimensional normal random variable with mean zero and variance V , where V is the covariance matrix defined in the theorem [see Rao, 1965, p. 108].

Using Slutsky's Theorem, Cramer [1946, p. 255],

$$n^{\frac{1}{2}} (\tilde{\beta} - \beta) \xrightarrow{L} N(0, \hat{\Sigma}_{XX}^{-1} V \hat{\Sigma}_{XX}^{-1}) . //$$

5. An estimator for the variance of the regression coefficients given non-normal error structure

We now consider the case where $x_t, (\varepsilon_t; u_t.)$, $t=1, 2, \dots, n$ have non-normal distributions. We call upon the following assumptions:

- i) $(y_t; x_t.)$ is a $1 \times (p+1)$ vector selected from a multivariate population with finite 4 δ moments, $t=1, 2, \dots, n$.

$(y_t; x_{t.})$ is independent of $(y_s; x_{s.})$ for $t \neq s$.

$x_{t.}$, $t=1, 2, \dots, n$ has mean vector zero and covariance matrix Σ_{xx} .

ii) $E\{(\epsilon_t; u_{t.}) | (y_t; x_{t.})\} = 0$, $t=1, 2, \dots, n$.

iii) The random variables $(\epsilon_t; u_{t.})$ are independent, identically distributed, have zero means, uniformly bounded $4+\delta$ moments, $\delta > 0$, and error covariance matrix

$$E \left\{ \begin{pmatrix} \epsilon_t \\ u_{t.} \end{pmatrix} (\epsilon_t; u_{t.}) \right\} = \Sigma = \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & D_{uu} \end{pmatrix}$$

for $t=1, 2, \dots, n$.

Theorem 4.D.2

Given assumptions i) - iii), then

$$n^{1/2} (\hat{\beta} - \beta) \xrightarrow{L} N(0, \Sigma_{xx}^{-1} A \Sigma_{xx}^{-1})$$

and

$$\frac{1}{n} (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} \tilde{A} (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1}$$

is a consistent estimator for the variance of $\tilde{\beta}$ where,

$$\tilde{\beta} = (\hat{M}_{XX} - \Lambda \hat{D}_{XX})^{-1} \hat{M}_{XY}, \quad (4.D.23)$$

$$A = nE (\hat{M}_{XV} + \hat{\Lambda}_{D_{XX}} \beta) (\hat{M}_{VX} + \beta' \hat{D}_{XX} \Lambda) , \quad (4.D.24)$$

$$\tilde{A} = \frac{1}{n-p} \sum_{n-p}^n \tilde{d}_{i.}' \tilde{d}_{i.} \quad (4.D.25)$$

$$\tilde{d}_{i.} = (\tilde{d}_{i1}, \tilde{d}_{i2}, \dots, \tilde{d}_{ip})$$

$$\tilde{d}_{ij} = x_{ij} \tilde{v}_i + \frac{\lambda_j}{1+\lambda_j} x_{ij}^2 \tilde{\beta}_j \quad (4.D.26)$$

$$\tilde{v}_i = y_i - x_{i.}' \tilde{\beta} \quad (4.D.27)$$

$$\hat{M}_{XV} = \frac{1}{n} X' v,$$

$\tilde{\beta}_j$ is the j^{th} vector element of β , $j=1, 2, \dots, p$.

Proof:

Our estimator is,

$$\tilde{\beta} = (\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}})^{-1} \hat{M}_{XY}$$

and from equation (4.D.17)

$$\tilde{\beta} - \beta = (\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}})^{-1} (\hat{M}_{XV} + \hat{\Lambda}_{D_{XX}} \beta) .$$

Now, $E\{(\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}})\} = \Sigma_{XX}$ and the variance of the elements of $(\hat{M}_{XX} - \hat{\Lambda}_{D_{XX}})$ are of order $\frac{1}{n}$. Hence, by Corollary 2.A.2

$$\hat{M}_{XX} - \Lambda \hat{D}_{XX} = \mathfrak{L}_{XX} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

The elements of $\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta$ are of order $n^{-1/2}$, hence

$$\tilde{\beta} - \beta = \mathfrak{L}_{XX}^{-1} (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta) + o_p\left(\frac{1}{n}\right). \quad (4.D.28)$$

From Theorem 2.B.1, it follows that the limiting distribution of $n^{1/2}(\tilde{\beta} - \beta)$ is the same as the limiting distribution of $n^{1/2} \mathfrak{L}_{XX}^{-1} \times (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta)$. Thus, let us investigate the limiting distribution of

$$n^{1/2} \mathfrak{L}_{XX}^{-1} (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta).$$

Let ρ be an arbitrary nonzero (px1) vector, and consider

$$n^{1/2} \rho' (\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta).$$

First, note that

$$(\hat{M}_{XV} + \Lambda \hat{D}_{XX} \beta) = \frac{1}{n} \sum_{i=1}^n \theta_i$$

where $\theta_i = x_i' v_i + \text{mdia}g(x_i' x_i) \Lambda \beta$ is the mean of n independent random variables by assumptions i) and iii). Second note that

$$E \left[\frac{1}{n} \sum_{i=1}^n \theta_i \right] = 0.$$

Third, note that

$$E[(\rho' \theta_i)^2] = \rho' E(\theta_i \theta_i') \rho,$$

is uniformly bounded, and

$$E[|\rho' \theta_i|^{2+\frac{1}{2}\delta}], \quad \delta > 0,$$

is uniformly bounded. This follows since the random variables $(y_t; x_t.)$ and $(\epsilon_t; u_t.)$ possess $4+\delta$ moments.

Now we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E[|\rho' \theta_i|^{2+\frac{1}{2}\delta}]}{\left\{ \sum_{i=1}^n E[(\rho' \theta_i)^2] \right\}^{\frac{1}{2}(2+\frac{1}{2}\delta)}} \\ &= \lim_{n \rightarrow \infty} \frac{\{n^{-1} \sum_{i=1}^n E[|\rho' \theta_i|^{2+\frac{1}{2}\delta}]\}}{n^{\frac{1}{4}\delta} \{ \rho' n^{-1} \sum_{i=1}^n E(\epsilon_i \theta_i') \rho \}^{1+\frac{1}{4}\delta}} = 0. \end{aligned}$$

By the Liapounov central limit theorem, Theorem 2.C.2, this gives

$$\frac{n^{-\frac{1}{2}} \sum_{i=1}^n \rho' \theta_i}{\left\{ \rho' E\left[\frac{1}{n} \sum_{i=1}^n (\theta_i \theta_i') \right] \rho \right\}^{\frac{1}{2}}} \xrightarrow{L} N(0,1).$$

Furthermore, by our assumptions

$$E \left[\frac{1}{n} \sum_{i=1}^n \theta_i \theta_i' \right] = O(1)$$

and thus by a result in Rao [1965, p. 102],

$$n^{-1/2} \sum_{i=1}^n \rho' \theta_i \xrightarrow{L} N(0, \rho' \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E(\theta_i \theta_i') \rho) .$$

We proceed to show that

$$A = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E(\theta_t \theta_t') .$$

The 1th diagonal element of

$$\frac{1}{n} \sum_{t=1}^n E(\theta_t \theta_t')$$

is

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n E \left(X_{i1} v_i + X_{i1}^2 \frac{\bar{\lambda}_1}{1+\lambda_1} \beta_1 \right)^2 \\ &= E(X_1^2 v^2) + 2 \frac{\lambda_1 \beta_1}{1+\lambda_1} E(X_1^3 v) + \left(\frac{\lambda_1}{1+\lambda_1} \right)^2 \beta_1^2 E(X_1^2) \end{aligned}$$

where X_1 denotes the 1th scalar random variable, $l=1, 2, \dots, p$.

The kl^{th} off-diagonal element of

$$\frac{1}{n} \sum_{i=1}^n E(\theta_i \theta_i')$$

is

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n E\left(X_{ik} v_t + X_{ik}^2 \frac{\lambda_k}{1+\lambda_k} \beta_k\right) \left(X_{il} v_i + X_{il}^2 \frac{\lambda_l}{1+\lambda_l} \beta_l\right) \\ &= E(X_{1k} X_{1k} v^2) + \frac{\lambda_k \beta_k}{1+\lambda_k} E(X_{tk}^2 X_{tl} v) \\ & \quad + \frac{\lambda_l \beta_l}{1+\lambda_l} E(X_{tl} X_{tk}^2 v) + \left(\frac{\lambda_l \beta_l}{1+\lambda_l}\right) \left(\frac{\lambda_k \beta_k}{1+\lambda_k}\right) E(X_k^2 X_l^2) . \end{aligned}$$

Using $E(\hat{M}_{XV} + \hat{\Delta}_{XX} \beta) = 0$, it easily follows that the diagonal and off-diagonal elements of A are the same as the diagonal and off-diagonal elements of

$$\frac{1}{n} \sum_{i=1}^n E(\theta_i \theta_i') .$$

Hence, by the multivariate central limit theorem, we now have

$$n^{-1/2} \sum_{i=1}^n \theta_i \xrightarrow{L} N(0, A) .$$

Finally, using equation (4.D.28), we obtain

$$n^{\frac{1}{2}} (\tilde{\beta} - \beta) \xrightarrow{L} N(0, \Sigma_{XX}^{-1} A \Sigma_{XX}^{-1}) .$$

We next prove that $(\hat{M}_{XX} - \hat{\Lambda} \hat{D}_{XX})^{-1} \tilde{A} (\hat{M}_{XX} - \hat{\Lambda} \hat{D}_{XX})^{-1}$ is a consistent estimator for the variance of β . Now, by our assumptions,

$$\begin{aligned} \tilde{v}_t &= Y_t - X_t \tilde{\beta} \\ &= v_t - X_t (\tilde{\beta} - \beta) . \end{aligned} \tag{4.D.29}$$

Also

$$\begin{aligned} \tilde{A} &= \frac{1}{n-p} \sum_{i=1}^n \tilde{d}_i \tilde{d}_i' \\ &= \frac{1}{n-p} \sum_{i=1}^n [X_i' \tilde{v}_i + \text{mdia}g(X_i' X_i) \Lambda \beta] [X_i' \tilde{v}_i + \text{mdia}g(X_i' X_i) \Lambda \beta]' . \end{aligned} \tag{4.D.30}$$

Substituting (4.D.29) into (4.D.30), we obtain

$$\begin{aligned} \tilde{A} &= \frac{1}{n-p} \sum_{i=1}^n [X_i' v_i + \text{mdia}g(X_i' X_i) \Lambda \beta]' \\ &\quad [X_i' v_i + \text{mdia}g(X_i' X_i) \Lambda \beta] + o_p(n^{-\frac{1}{2}}) , \\ &= \hat{A} + o_p(n^{-\frac{1}{2}}) \end{aligned}$$

where

$$\frac{1}{n-p} \sum_{i=1}^n [x'_{i.} v_i + \text{mdia}g (x'_{i.} x_{i.}) \wedge \beta]'$$

$$[x'_{i.} v_i + \text{mdia}g (x'_{i.} x_{i.}) \wedge \beta] .$$

The elements of the matrix $\frac{n-p}{n} \hat{A}$ have means equal to the corresponding elements in the matrix A . Furthermore, by assumptions i) and iii), the elements of $\frac{n-p}{n} \hat{A}$ are the means of independent random variables with finite $1 + \frac{1}{2}\delta$ moments. It follows by Markov's weak law of large numbers, Parzen [1960, p. 418] that

$$\frac{1}{n-p} \sum_{i=1}^n \tilde{d}'_{i.} \tilde{d}_{i.} \xrightarrow{P} A . \quad //$$

6. A modified estimator

We modify the estimator β in order to guarantee the existence of the first two moments.

Consider

$$\tilde{\beta}_2 = \hat{H}_{XX}^{-1} \hat{M}_{XY} , \quad (4.D.23)$$

where

$$\hat{H}_{XX} = \hat{M}_{XX} - \hat{\Lambda}_{XX}^D \quad \text{if } \hat{\gamma} \geq 1 + \frac{1}{n} ,$$

$$= \hat{M}_{XX} - (\hat{\gamma} - \frac{1}{n}) \hat{\Delta D}_{XX} \quad \text{if } \hat{\gamma} < 1 + \frac{1}{n} .$$

$\hat{\gamma}$ is the smallest root of

$$|\hat{M}_{XX} - \gamma(\hat{\Delta D}_{XX})| = 0 .$$

Theorem 4.D.3

Given assumptions i)-v) of Section 4.D.2, the i^{th} moment of the estimator

$$\tilde{\beta}_2 = \hat{H}_{XX}^{-1} \hat{M}_{XY}$$

is of order one.

Proof:

The i^{th} element of $\tilde{\beta}_2$ is given by

$$\begin{aligned} \tilde{\beta}_{2i} &= \sum_{j=1}^P h_{ij} m_j \\ &= \sum_{j=1}^P |\hat{H}_{XX}|^{-1} \text{Cof}(h_{ij}) m_j \end{aligned}$$

where $\text{Cof}(h_{ij})$ is the signed cofactor of h_{ij} , h^{ij} is the ij^{th} element of \hat{H}_{XX}^{-1} and m_j is the j^{th} element of \hat{M}_{XY} .

Now for $n > N_0$,

$$|\hat{H}_{XX}| \geq |\frac{1}{n} \hat{\Delta D}_{XX}| .$$

Let

$$Y(\hat{\theta}) = |\Lambda \hat{D}_{XX}|^{-r} \left| \sum_{j=1}^p \text{Cof}(h_{ij}) m_j \right|^r$$

where $\hat{\theta}$ denotes the column vector obtained by arranging the p columns of \hat{M}_{XX} and \hat{M}_{XY} in a single column.

The elements of $\hat{\theta}$ are distributed with mean θ and variances of order n^{-1} . In order to establish the conditions of Theorem 2.D.2 we note that, for sufficiently large n , the elements of $\tilde{\beta}_2$ are continuous differentiable functions of θ for all θ in a bounded span at S containing the true value θ .

Now,

$$|\tilde{\beta}_{2i}|^r \leq n^{rp} Y(\hat{\theta}) .$$

$|\hat{D}_{XX}|$ may be expressed as a multiple of its diagonal elements

$$|\hat{D}_{XX}| = \prod_{i=1}^p \frac{X_i' X_i}{n} .$$

The expected value of $\frac{X_i' X_i}{n}$ is $\sigma_{X_i}^2$ where

$$\sigma_{X_i}^2 = \sigma_{x_i}^2 + \sigma_{u_i}^2 .$$

Defining $z_i = \frac{x_i' x_i}{n}$, it is easily seen that $\frac{z_i}{\sigma_{X_i}^2}$ is distributed as

a χ^2 with p -degrees of freedom for $i=1, 2, \dots, p$.

Now,

$$\begin{aligned} \int |\hat{D}_{XX}|^{-4r} dF_n(\hat{\theta}) &= \prod_{i=1}^p \int (\Lambda_i z_i)^{-4r} dF_n(\hat{\theta}) \\ &= \prod_{i=1}^p \left(\frac{1}{2}\right)^{2r} \frac{\Gamma(\frac{p-4r}{2})}{\Gamma(\frac{p}{2})} (\sigma_{X_i}^2)^{4r} . \\ &= O(1) . \end{aligned}$$

Since

$$\begin{aligned} E(\hat{\theta} - \theta)^{2r} &= \int \|\hat{\theta} - \theta\|^{2r} dF_n(\hat{\theta}) \\ &= O(n^{-r}) \quad r=1, 2, \dots . \end{aligned}$$

We have

$$\int \left| \sum_{j=1}^p \text{Cof}(h_{ij}) m_j \right|^{4r} = O(1) .$$

Using the Cauchy-Schwartz Inequality,

$$\int |Y(\hat{\theta})|^2 dF_n(\hat{\theta}) \leq \left[\int |\hat{D}_{XX}|^{-4r} dF_n(\hat{\theta}) \right]^{\frac{1}{2}} \left[\int \left| \sum_{j=1}^p \text{Cof}(h_{ij}) m_j \right|^{4r} \right]^{\frac{1}{2}} \\ = o(1) .$$

Hence the conditions of Theorem 2.D.2 are satisfied. We have obtained the required result. //

E. An Example

Winakor [1973], in a study of textile expenditure by households, used a stratified area-sample of Central Iowa households. In this example, we will consider data from 231 clusters in urban areas.

We postulate that textile expenditure of a household, as measured in dollars, is a function of the three variables

- a) income--income of the family in hundreds of dollars for the family units
- b) family size--the number of people in a family
- c) moving--a zero one variable indicating whether the family had moved within the last year; zero indicates that the family unit had not moved and one indicates that it had moved.

To obtain residuals with constant variance, the data was transformed using a natural log transformation. It is assumed that

$$y_{ij} = \log_e (\text{Expenditure} + 5) ,$$

$$x_{ij1} = \log_e (\text{Income} + 5) ,$$

$$x_{ij2} = \log_e (\text{Family Size}) ,$$

and

$$x_{ij3} = \text{Moving} .$$

Income and family size are variables assumed to be measured with error. The moving variable is zero-one and it is assumed to be measured without error. It is assumed that the response error in income and family size are mutually independent.

The response variance and total variance for $\log_e (\text{Income} + 5)$ and $\log_e (\text{Family Size})$ were estimated from the U.S. Census publication (U.S. Bureau of the Census [1972, pages 50 and 95]). In the Census report, a replication study was designed to investigate some of the effects of using different-data-gathering techniques. Two samples were included in the replication study.

a) In Sample I, composed of 5,000 housing units, enumerators were sent to the housing units and the persons at the units responded by means of a direct interview.

b) In Sample II, composed of 1,000 housing units, questionnaires identical with the census questionnaires were mailed to the housing units. The householders were asked to mail in the completed questionnaires within 3 days. Nonresponse and inconsistencies were followed up by enumerators.

In both samples, the householders were asked to make the second response within 2 or 3 weeks of the census. In both samples, the reinterview questionnaire was identical to the census questionnaire. The data thus obtained was classified in a two-way contingency table, one way for census classification and the other way for reinterview classification.

Table 1 shows the data classified by response to the question on number of children in both the census and reinterview for Sample I. The data pertains to all women 14 years old and over who had been married at the time of the census. Table 2 shows the data classified by reported income. Reported income is the sum of wages and salary for self employment, and other income, using Census definitions.

Note that the Census data used to estimate the total variance and response variance for the Winakor data is not the most appropriate, however, it was the only available source of information to estimate these variances. There are several reasons for this. First, we utilized the number of children ever born to married women fourteen years old and over, to estimate total variance and response variance of reported family size. These estimates are not

Table 1. Estimates of differences in reporting number of children ever born to females 14 years old and over, ever married, Sample I. Data Source: U.S. Bureau of the Census [1972, p. 95]

Reinterview classification	Census classification					
	Females, 14 years old and over, marital status reported in census					
	Total	None	1	2	3	4
Total	36,127	5,437	6,315	9,737	6,011	3,091
None	5,798	5,207	243	148	- ^a	41
1	6,696	149	6,237	258	51	-
2	3,706	40	352	9,019	234	60
3	6,010	-	43	224	5,622	121
4	3,062	40	-	49	103	2,869
5	1,671	-	-	39	-	-
6	1,142	-	-	-	-	-
7	849	-	-	-	-	-
8	486	-	-	-	-	-
9	330	-	-	-	-	-
10	71	-	-	-	-	-
11	83	-	40	-	-	-
12 or more	222	-	-	-	-	-

^a - represents zero.

5	6	7	8	9	10	11	12 or more
1,724	1,195	780	570	246	117	82	222
119	40	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
1,474	99	60	-	-	-	-	-
-	1,001	-	141	-	-	-	-
89	55	677	28	-	-	-	-
43	-	43	401	-	-	-	-
-	-	-	-	246	46	39	-
-	-	-	-	-	71	-	-
-	-	-	-	-	-	43	-
-	-	-	-	-	-	-	222

Table 2. Estimates of differences in reporting total income of males 14 years old and over, Sample I. Data Source: U.S. Bureau of the Census [1972, p. 95]

Reinterview classification	Census classification						
	Males, 14 years old and over, total income reported in census						
	Total	None	\$1 to \$499	\$500 to \$999	\$1000 to \$1499	\$1500 to \$1999	\$2000 to \$2499
Total	32435	813	1802	1558	1485	1257	1441
None	970	519	152	73	49	74	36
\$1 to \$499	1707	124	1167	237	86	14	-
\$500 to \$999	1520	64	244	904	152	48	21
\$1000 to \$1499	1311	11	79	122	660	213	58
\$1500 to \$1999	1379	27	24	68	259	623	166
\$2000 to \$2499	1651	32	68	73	98	141	710
\$2500 to \$2999	1377	12	8	24	27	32	215
\$3000 to \$3499	1507	9	9	10	66	20	75
\$3500 to \$3999	1622	- ^a	4	11	23	47	42
\$4000 to \$4499	2428	14	24	-	12	9	68
\$4500 to \$4999	2400	-	-	-	12	-	8
\$5000 to \$5999	5127	-	7	10	17	-	21
\$6000 to \$6999	3287	-	9	-	7	7	-
\$7000 to \$9999	4118	-	7	12	9	10	11
\$10000 and over	1971	-	-	13	8	11	11

^a - represents zero.

\$2500 to \$2999	\$3000 to \$3499	\$3500 to \$3999	\$4000 tp \$4499	\$4500 to \$4999	\$5000 to \$5999	\$6000 to \$6999	\$7000 to \$9999	\$10000 and over
1304	1664	1571	2392	2275	5287	3517	4193	1877
-	7	-	22	-	20	8	10	-
11	-	-	30	15	10	-	-	13
10	40	-	11	-	-	-	16	9
53	56	17	8	11	-	-	13	8
82	23	11	16	-	19	10	22	22
188	160	40	94	8	10	17	-	12
566	220	36	108	31	36	11	40	9
125	663	183	150	44	108	8	28	9
50	285	807	153	100	72	-	29	-
50	81	272	1168	338	216	89	68	20
33	54	38	367	1192	566	67	64	-
107	75	46	203	398	3324	663	203	52
13	-	23	35	83	544	2113	366	87
-	-	9	28	46	205	452	3092	239
13	-	30	-	11	156	79	242	1397

the best, since, family size in the Winakor study comprises the two parents and their living dependents, whereas, in the Census, only the number of children ever born is reported. Hence, there is reason to believe that the number of children reported by the Census figures would be higher than the number of children reported in the Winakor study. This would imply that the estimate of response variance obtained using the Census data would overestimate the response variance in the Winakor study. Second, we used the Census estimate of total income of males fourteen years old and over to estimate total variance and response variance of family income. Again, these estimates are not the best since family income is comprised of the sum of the income of each member of the family.

For the data presented in Tables 1 and 2, let m be the number of classifications. Thus the interview-reinterview table is an $m \times m$ array where the ij^{th} cell, $i, j=1, 2, \dots, m$; the number of persons who were classified in the i^{th} category in the census and in the j^{th} category in the reinterview. Let the number of persons classified in the ij^{th} cell be f_{ij} . We denote by w_i , $i=1, 2, \dots, m$, the value associated with the i^{th} classification. For example, in Table 1, w_i is the number of children ever born to females 14 years old and over, ever married. If twelve or more children were born, we set w_i equal to fifteen. In Table 2, w_i is the mid-point for each class interval for total income reported. For total income over \$10,000, we set w_i equal to \$14,000.

If computations are to be made from Table 1, define

$$s_i = \log_e (w_i + 2) \quad i=1, 2, \dots, m;$$

where w_i = number of children reported in the i^{th} classification.

If computations are to be made from Table 2, or define $s_i = \log_e (w_i/100 + 5)$ where w_i = total income value used for the i^{th} classification.

Using the data in Table 1 or Table 2, the estimated total variance is estimated by

$$\hat{\sigma}_T^2 = \frac{1}{2} (\hat{\sigma}_R^2 + \hat{\sigma}_C^2)$$

where

$$\hat{\sigma}_R^2 = \frac{1}{\sum_{j=1}^m f_{.j}} \left[\sum_{j=1}^m f_{.j} s_j^2 - \frac{(\sum_{j=1}^m f_{.j} s_j)^2}{\sum_{j=1}^m f_{.j}} \right]$$

$$\hat{\sigma}_C^2 = \frac{1}{\sum_{i=1}^m f_{i.}} \left[\sum_{i=1}^m f_{i.} s_i^2 - \frac{(\sum_{i=1}^m f_{i.} s_i)^2}{\sum_{i=1}^m f_{i.}} \right]$$

$$f_{i.} = \sum_{j=1}^m f_{ij}$$

$$f_{.j} = \sum_{i=1}^m f_{ij}$$

The response error variance is directly estimated by

$$\hat{\sigma}_u^2 = \frac{1}{2} \frac{\sum_{i=1}^m \sum_{j=1}^m f_{ij} (s_i - s_j)^2}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}} .$$

The estimated total variances for $\log_e (\text{income} + 5)$ and $\log_e (\text{Family Size})$ were 0.6227 and 0.1977 respectively.

The estimated response error variances for $\log_e (\text{Income} + 5)$ and $\log_e (\text{Family Size})$ were 0.0730 and 0.0096 respectively. Thus, the ratio of response error variances to true variances for $\log_e (\text{Income} + 5)$ was 0.1328 and for $\log_e (\text{Family Size})$ was 0.0510.

We assume the following model for the data

$$y_{ij} = \beta_0 + \sum_{k=1}^3 \beta_k x_{ijk} \quad (4.E.1)$$

where

$i=1, 2, \dots, 231$ --the number of clusters in urban areas,

$j=1, 2, \dots, n_i$ --the number of secondary units in the i^{th} cluster.

In this example, we have 630 families in the 231 clusters.

In matrix notation model (4.E.1) may be written as,

$$y = x \beta \quad (4.E.2)$$

where y is an 630×1 vector, x is an 630×4 matrix and β is 4×1 . The β 's are the unknown parameters that we wish to estimate. We observe Y , which is the sum of y and a measurement error e , and X which is the sum of x and a measurement error u . In matrix notation corresponding to (4.E.2), we have

$$\begin{aligned} Y &= y + e \\ X &= x + u \end{aligned} \tag{4.E.3}$$

Three models for the data and associated methods are considered and illustrated.

A. In this model, we assume that x is an 630×4 matrix of fixed constants, $\epsilon = (e; u)$ is an 630×5 matrix of random variables whose rows are independently and identically distributed with mean zero and covariance matrix Σ . That is,

$$(e_t; u_t) \sim \text{NID}(0, \Sigma)$$

where $(e_t; u_t)$ denotes the t^{th} row of the matrix ϵ and

$$\Sigma = \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \Sigma_{uu} \end{pmatrix},$$

with Σ_{uu} being a 4×4 matrix with off-diagonal elements equal to zero. It is assumed that Σ_{uu} is known. Also we assume that $(e; u)$ is independent of $(y; x)$.

A moment estimator of β is

$$\hat{\beta} = [X'X - 626 \hat{\Sigma}_{uu}]^{-1} X'Y .$$

B. We assume model (4.D.1) with assumptions i) - iv) of subsection (4.D.2). That is, we are given the ratio of the error variance to the total variance. We compute the estimator β defined by Equation (4.D.7).

C. We use the ordinary least squares regression estimator

$$\hat{\beta}_{OLS},$$

$$\hat{\beta}_{O.L.S.} = (X'X)^{-1} X'Y .$$

The computations involved in those three procedures are given in Tables 4 through 9. Table 3 summarizes the results for these three methods. From Table 3, one immediately observes that, compared with the least-squares estimates, the errors-in-variables estimated coefficients are larger using methods A and B for $\text{Log}_e (\text{Income} + 5)$ (0.5941 vs. 0.9374, 0.6758), $\text{Log}_e (\text{Size})$ (0.3668 vs. 0.3799, 0.3835) and Moving (0.3905 vs. 0.4735, 0.4111). The estimated standard errors for each of the errors-in-variables estimators are larger than the estimated standard errors of the O.L.S. estimator. The regression coefficients for $\text{Log}_e (\text{Income} + 5)$ using method A is larger than the regression coefficient for $\text{Log}_e (\text{Income} + 5)$ using method B. For all the regression coefficients of interest and all

the methods used, their associated t values are all significant at the five percent level. Both the regression coefficient and its associated standard error for $\log_e (\text{Income} + 5)$ and Moving are larger in method A than in method B. This occurs since the total variance of $\log_e (\text{Income} + 5)$ for the Winakor data (.2058) is smaller than the total variance of $\log_e (\text{Income} + 5)$ for the U.S. Census data (0.6227).

We have presented two methods for estimating the regression coefficients and their associated estimated standard errors given an errors-in-variables model. Our method for estimating the variances of the regression coefficients does not require the assumption that the errors of measurement are normally distributed. We only need the measurement errors to have finite $4 + \delta$, $\delta > 0$, moments.

Table 3. Summary of results

	$\text{Log}_e (\text{inc.}+5)$	$\text{Log}_e (\text{Size})$	Moving
Errors-in-variables regression method A	0.9374	0.3799	0.4735
Estimated standard errors	0.1797	0.1138	0.1319
t	5.216*	3.340*	3.590*
Errors-in-variables regression method B	0.6761	0.3835	0.4112
Estimated standard errors	0.1200	0.1118	0.1283
t	5.635*	3.431*	3.205*
Least squares regression method C	0.5941	0.3668	0.3305
Estimated standard errors	0.1076	0.1059	0.1267
t	5.519*	3.464*	3.081*

*Significant at the 5% level.

Table 4. Calculation of $\hat{\beta}$ for Method A

$$\hat{\beta} = (X'X - 626 \Sigma_{uu})^{-1} X'Y$$

where the rs^{th} element of $X'X$ is

$$\sum_{i=1}^{231} \sum_{j=1}^{n_i} X_{ijr} X_{ijs}$$

and the r^{th} element of $X'Y$ is

$$\sum_{i=1}^{231} \sum_{j=1}^{n_i} X_{ijr} Y_{ij}.$$

The matrices of interest are

$$X'X = \begin{bmatrix} 630.0 & 2939.8 & 810.4 & 102.0 \\ 2939.8 & 13848.2 & 3785.6 & 455.5 \\ 810.4 & 3785.6 & 1148.8 & 125.5 \\ 102.0 & 455.5 & 125.5 & 102.0 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 2229.4 \\ 10474.0 \\ 290.9 \\ 380.1 \end{bmatrix}$$

$$\Sigma_{uu} = \text{diag} (0.0000, 0.0730, 0.0096, 0.0000)$$

Table 4. (continued)

$$\hat{\beta} = (X'X - 626 \mathbb{1}_{uu})^{-1} X'Y$$

$$X'X - 626 \mathbb{1}_{uu} = \begin{bmatrix} 630.0 & 2939.8 & 810.4 & 102.0 \\ 2939.8 & 13802.5 & 3785.6 & 455.5 \\ 810.4 & 3785.6 & 1148.8 & 125.5 \\ 102.0 & 455.5 & 125.5 & 102.0 \end{bmatrix}$$

Therefore

$$\hat{\beta} = \begin{pmatrix} -1.401 \\ 0.937 \\ 0.380 \\ 0.473 \end{pmatrix}.$$

Table 5. Calculation of $\hat{\text{Var}}(\hat{\beta})$ for Method A

$$\hat{\text{Var}}(\hat{\beta}) = (X'X - 626 \text{ } \hat{\text{I}}_{uu})^{-1} \hat{G} (X'X - 626 \text{ } \hat{\text{I}}_{uu})^{-1} .$$

The rs^{th} element of \hat{G} is

$$\hat{g}_{rs} = \sum_{i=1}^{231} \frac{n_i}{n_i - 1} (\hat{d}_{i \cdot r} - \bar{\hat{d}}_{\cdot \cdot r})(\hat{d}_{i \cdot s} - \bar{\hat{d}}_{\cdot \cdot s})$$

where

$$\hat{d}_{ijr} = x_{ijr} \hat{e}_{ij} \quad r=1, 2, \dots, 4$$

$$\hat{d}_{i \cdot r} = \sum_{j=1}^{n_i} \hat{d}_{ijr}$$

$$\bar{\hat{d}}_{\cdot \cdot r} = \frac{1}{231} \sum_{i=1}^{231} \hat{d}_{i \cdot r}$$

$$\hat{e}_{ij} = y_{ij} - \hat{\beta}_0 - \sum_{k=1}^3 x_{ijk} \hat{\beta}_k .$$

Calculating

$$\hat{G} = \begin{bmatrix} 990.4 & 4670.9 & 1244.9 & 168.4 \\ 4670.9 & 22231.9 & 5843.7 & 770.3 \\ 1244.9 & 5843.7 & 1691.3 & 206.6 \\ 168.4 & 770.3 & 206.6 & 141.1 \end{bmatrix}$$

Table 5. (continued)

$$\hat{\text{Var}}(\hat{\beta}) = (X'X - 626 \mathbb{I}_{uu})^{-1} \hat{G} (X'X - 626 \mathbb{I}_{uu})^{-1} .$$

we obtain

$$\hat{\text{Var}}(\hat{\beta}) = \begin{bmatrix} 0.6694 & -0.1444 & 0.0067 & -0.0389 \\ -0.1444 & 0.0323 & -0.0051 & 0.0078 \\ 0.0067 & -0.0051 & 0.0123 & -0.0001 \\ -0.0389 & 0.0078 & -0.0001 & 0.0174 \end{bmatrix} .$$

Table 6. Calculation of β for Method B

$$\tilde{\beta} = [X'X - 630 \Lambda S_{XX}]^{-1} X'Y$$

where

$$X'X = \begin{bmatrix} 630.0 & 2939.8 & 810.4 & 102.0 \\ 2939.8 & 13848.2 & 3785.6 & 455.5 \\ 810.4 & 3785.6 & 1148.8 & 125.5 \\ 102.0 & 455.5 & 125.5 & 102.0 \end{bmatrix}$$

$$\Lambda = \text{diag} (0.0000, 0.1172, 0.0485, 0.0000)$$

$$S_{XX} = \text{diag} (0.0000, 0.2055, 0.1692, 0.1359) .$$

The k^{th} diagonal element of S_{XX} is

$$S_{XX}(k) = \frac{1}{629} \sum_{i=1}^{231} \sum_{j=1}^{n_i} (x_{ijk} - \bar{x}_{..k})^2$$

where

$$\bar{x}_{..k} = \frac{1}{630} \sum_{i=1}^{231} \sum_{j=1}^{n_i} x_{ijk} .$$

Table 6. (continued)

$$\tilde{\beta} = [X'X - 630 \quad S_{XX}]^{-1} X'Y$$

Now

$$X'Y = \begin{bmatrix} 2229.4 \\ 10474.0 \\ 290.9 \\ 380.1 \end{bmatrix},$$

and

$$X'X - 630 \wedge S_{XX} = \begin{bmatrix} 630.0 & 2939.8 & 810.4 & 102.0 \\ 2939.8 & 13833.0 & 3785.6 & 455.5 \\ 810.4 & 3785.6 & 1143.6 & 125.5 \\ 102.0 & 455.5 & 125.5 & 102.0 \end{bmatrix}$$

Therefore

$$\tilde{\beta} = \begin{pmatrix} -0.176 \\ 0.676 \\ 0.383 \\ 0.411 \end{pmatrix}.$$

Table 7. Calculations of $\tilde{\text{Var}}(\tilde{\beta})$ for Method B

$$\tilde{\text{Var}}(\tilde{\beta}) = [X'X - 630 \Lambda s_{XX}]^{-1} \tilde{G} [X'X - 630 \Lambda s_{XX}]^{-1}$$

The r_s^{th} element of \tilde{G} is

$$\tilde{g}_{rs} = \frac{630}{626} \sum_{i=1}^{231} \sum_{j=1}^{n_i} \tilde{d}_{ijr} \tilde{d}_{ijs}$$

where

$$\tilde{d}_{ijr} = x_{ijr} \tilde{v}_{ij} + \frac{\lambda_r}{1+\lambda_r} x_{ijr}^2 \tilde{\beta}_r, \quad r=1, 2, \dots, 4$$

$$\tilde{v}_{ij} = y_{ij} + 0.176 - 0.676 x_{ij1} - 0.383 x_{ij2} - 0.411 x_{ij3}.$$

We have

$$[X'X - 630 \Lambda s_{XX}]^{-1} = \begin{bmatrix} 0.2180 & -0.0427 & -0.0118 & -0.0129 \\ -0.0427 & 0.0022 & -0.0002 & 0.0022 \\ -0.0118 & -0.0002 & 0.0099 & 0.0006 \\ -0.0129 & 0.0022 & 0.0006 & 0.0122 \end{bmatrix}$$

and

$$\tilde{G} = \begin{bmatrix} 749.9 & 3514.9 & 937.8 & 138.9 \\ 3514.9 & 16652.8 & 4384.8 & 625.1 \\ 937.8 & 4384.8 & 1298.3 & 167.7 \\ 138.9 & 625.1 & 167.7 & 138.9 \end{bmatrix}.$$

Table 7. (continued)

$$\tilde{\text{Var}}(\tilde{\beta}) = [x'x - 630 \wedge s_{xx}]^{-1} \tilde{G}[x'x - 630 \wedge s_{xx}]^{-1}$$

Hence,

$$\tilde{\text{Var}}(\tilde{\beta}) = \begin{bmatrix} 0.3182 & -0.0651 & -0.0081 & -0.0192 \\ -0.0651 & 0.0144 & -0.0018 & 0.0035 \\ -0.0081 & -0.0018 & 0.0125 & 0.0005 \\ -0.0192 & 0.0035 & 0.0005 & 0.0164 \end{bmatrix}.$$

Table 8. Calculation of $\hat{\beta}_{O.L.S.}$ for Method C

$$\hat{\beta}_{O.L.S.} = (X'X)^{-1} X'Y$$

where the rs^{th} element of $X'X$ is

$$\sum_{i=1}^{231} \sum_{j=1}^{n_i} x_{ijr} x_{ijs}$$

and the r^{th} element of $X'Y$ is

$$\sum_{i=1}^{231} \sum_{j=1}^{n_i} x_{ijr} y_{ij} .$$

The numerical values for the elements of $X'X$ and $X'Y$ were given in Table 4

$$\hat{\beta}_{O.L.S.} = \begin{pmatrix} 0.2311 \\ 0.5941 \\ 0.3668 \\ 0.3905 \end{pmatrix} .$$

Table 9. Calculation of $\hat{\text{Var}}(\hat{\beta}_{\text{O.L.S.}})$ for Method C

$$\hat{\text{Var}}(\hat{\beta}_{\text{O.L.S.}}) = (X'X)^{-1} \hat{G}_{\text{O.L.S.}} (X'X)^{-1}$$

Where the r_s^{th} element of $\hat{G}_{\text{O.L.S.}}$ is

$$\hat{g}_{rs} = \sum_{i=1}^{231} \frac{n_i}{n_i - 1} (\hat{d}_{i \cdot r} - \bar{\hat{d}}_{\cdot \cdot r})(\hat{d}_{i \cdot s} - \bar{\hat{d}}_{\cdot \cdot s})$$

$$\hat{d}_{ijr} = x_{ijr} \hat{e}_{ij} \quad r=1, 2, \dots, 4$$

$$\hat{d}_{i \cdot r} = \sum_{j=1}^{n_i} \hat{d}_{ijr}$$

$$\bar{\hat{d}}_{\cdot \cdot r} = \frac{1}{231} \sum_{i=1}^{231} \hat{d}_{i \cdot r}$$

$$\hat{e}_{ij} = y_{ij} - 0.2311 - 0.5941 x_{ij1} - 0.3668 x_{ij2} - 0.3905 x_{ij3} .$$

Calculating

$$\hat{G}_{\text{O.L.S.}} = \begin{bmatrix} 956.5 & 4525.3 & 1206.1 & 161.2 \\ 4525.3 & 21588.4 & 5686.5 & 742.9 \\ 1206.1 & 5686.5 & 1645.2 & 198.6 \\ 161.2 & 742.9 & 198.6 & 138.6 \end{bmatrix}$$

Table 9. (continued)

$$\hat{\text{Var}} (\hat{\beta}_{\text{O.L.S.}}) = (X'X)^{-1} \hat{G}_{\text{O.L.S.}} (X'X)^{-1}$$

we obtain

$$\hat{\text{Var}} (\hat{\beta}_{\text{O.L.S.}}) = \begin{bmatrix} 0.2455 & -0.0509 & -0.0046 & -0.0167 \\ -0.0509 & 0.0116 & -0.0022 & 0.0030 \\ -0.0046 & -0.0022 & 0.0112 & 0.0004 \\ -0.0167 & 0.0030 & 0.0004 & 0.0161 \end{bmatrix}$$

V. A MONTE CARLO STUDY

In Chapter 3, we investigated the limiting behavior of the estimator of the finite population regression coefficient, given that the elements of the finite population were selected from a multivariate population with finite fourth moments. A Monte-Carlo study was undertaken to investigate the performance of the estimators under small sample conditions. The small sample properties of two regression estimators were studied. These regression estimators were:

- 1) The O.L.S. regression estimator and
- 2) an errors-in-variables regression estimator.

The data used for this investigation were those used by Frankel [1971] and were collected by the U.S. Bureau of the Census in the March 1967 Current Population Survey [1963]. The finite population consisted of 45,737 observations grouped in 3240 primary units. Two sample designs were used in this investigation. In sample design I the original 3240 primary units in the population were divided into 6 strata containing 540 primary units each. In sample design II, the 3240 primary units were divided into 12 strata, each of size 270 primary units. This stratification was carried out by splitting each of the 6 strata used in design I into two strata. In sample designs I and II, two primary sampling units were selected without replacement from each stratum of the population. The data was stored on a tape. Each individual element stored on this tape

was identified by a household number and a p.s.u. code. The p.s.u. numbers were ordered from 1 to 3240. All the elements associated with a specific p.s.u. were grouped together within the strata defined by the position of the p.s.u. on the sequence. In the case of the 6 strata design, the first 540 p.s.u. made up stratum I, the second 540 p.s.u. made up stratum II, etc. In the case of the 12 strata design, each stratum was arranged in a sequence of 270 p.s.u. Each of the two sampling designs called for the selection of two primary sampling units from each stratum of the population. A computer program was written to select the two primary sampling units using a simple random without replacement sampling scheme. For sample design I, 6 independent pairs of random numbers were generated. Each element of the pair was generated by a uniform (0,1) random number generator. The elements of each pair were multiplied by 540 and the product was truncated. For sample design II, 12 independent pairs of random numbers were generated, with each element of the pair generated by a uniform (0,1) random number generator. The elements of each pair were multiplied by 270. Two hundred independent samples were selected in this manner for each sampling design.

The dependent variable of interest was log Income of the household head and the independent variables were age, age squared and education. To insure that the matrix of sums of squares and products of the independent variables was nonsingular, the independent variables were coded as: Age-43, $(\text{Age}-43)^2$ -70 and Education-12. We let X_{ijk_r} ($r=1, 2, \dots, 4$) denote the value of the r^{th}

independent variable and Y_{ijk} the value of the dependent variable for the k^{th} element ($k=1, 2, \dots, M_{ij}$) in the j^{th} cluster ($j=1, 2, \dots, N_i$) of the i^{th} stratum ($i=1, 2, \dots, L$) of the population. Similarly, x_{ijk_r} denotes the value of the r^{th} independent ($r=1, 2, \dots, 4$) variable and y_{ijk} the value of the dependent variable for the k^{th} element ($k=1, 2, \dots, M_{ij}$) in the j^{th} primary ($j=1, 2$) of the i^{th} stratum ($i=1, 2, \dots, L$) of a selected sample. In addition, we define

$$X_{ijk1} = 1 \text{ and } x_{ijk1} = 1$$

for all i, j and k .

We investigated the sampling behavior of two sets of statistics, those associated with the O.L.S. estimator discussed in Fuller [1974, p. 12] and those associated with the errors-in-variables procedure discussed in Fuller [1974, p. 19]. Fuller [1974] extended Theorem 3.E.1 to the regression coefficient estimated from two-stage stratified samples. For the case of errors-in-variables, Fuller [1974] assumed that the response errors are independent between secondary units (clusters in our case) within the same primary unit (stratum) as well as between secondary units in different primary units.

For ordinary least squares, the population parameter is

$$B = Q_N^{-1} H_N$$

where Q_N is a (4×4) matrix with rs^{th} element

$$q_{Nrs} = \sum_{i=1}^L \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} x_{ijk r} x_{ijk s}$$

and H_N is a (4×1) vector with s^{th} element

$$h_{Ns} = \sum_{i=1}^L \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} x_{ijk s} y_{ijk} ,$$

$r, s = 1, 2, \dots, 4$; $L = 6, 12$; $N_i = 540, 270$.

For each selected sample, the sample regression estimator is

$$b = Q_n^{-1} H_n$$

where the rs^{th} element of Q_n is defined as

$$q_{nrs} = \sum_{i=1}^L \sum_{j=1}^2 \sum_{k=1}^{M_{ij}} x_{ijk r} x_{ijk s}$$

and the s^{th} element of H_n is defined as

$$h_{ns} = \sum_{i=1}^L \sum_{j=1}^2 \sum_{k=1}^{M_{ij}} x_{ijk s} y_{ijk} ,$$

$r, s = 1, 2, \dots, 4$; $L = 6, 12$.

The consistent estimator for the variance-covariance matrix of b is

$$\hat{V} = Q_n^{-1} \hat{G}_n Q_n^{-1} ,$$

where the rs^{th} element of \hat{G}_n is

$$\hat{g}_{nrs} = 2 \sum_{i=1}^L \sum_{j=1}^2 (\hat{d}_{ij \cdot r} - \hat{d}_{i \cdot \cdot r})(\hat{d}_{ij \cdot s} - \hat{d}_{i \cdot \cdot s})$$

and

$$\hat{d}_{ijk r} = x_{ijk r} \hat{e}_{ijk}$$

$$\hat{e}_{ijk} = y_{ijk} - \sum_{r=1}^4 b_r x_{ijk r}$$

$$\hat{d}_{ij \cdot r} = \sum_{k=1}^{M_{ij}} \hat{d}_{ijk r}$$

$$\hat{d}_{i \cdot \cdot r} = \frac{1}{2} \sum_{j=1}^2 \hat{d}_{ij \cdot r},$$

$i=1, 2, \dots, L$; $j=1, 2$; $k=1, 2, \dots, M_{ij}$; $r,s=1, 2, \dots, 4$.

Also of interest is the "t-statistics"

$$t(b_r) = \frac{b_r - B_r}{s(b_r)}, \quad r=1, 2, \dots, 4,$$

where $s(b_r)$ is the square root of the r^{th} diagonal element of the 4×4 matrix \hat{V} .

For the errors-in-variables model, age and education were observed subject to response error. Response variances, for Age-43,

$(\text{Age}-43)^2$ -70 and Education-12, were assumed to be 0.3, 91.0 and 3.0 respectively. It was assumed that the response errors in the three variables were uncorrelated and that the response error of variance was uncorrelated with that of age and education.

The population parameter for the errors-in-variables is

$$B(e) = Q_N^{-1}(e) H_N(e)$$

where the rs^{th} element of the 4×4 matrix is

$$q_{Nrs}(e) = \sum_{i=1}^L \sum_{j=1}^{N_i} \sum_{R=1}^{M_{ij}} (X_{ijkR} X_{ijks} - \sigma_{u_r u_s})$$

the s^{th} element of $H_N(e)$ is

$$h_{Ns}(e) = \sum_{i=1}^L \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} X_{ijks} Y_{ijk} ,$$

and $\sigma_{u_r u_s}$ is the rs^{th} element of the 4×4 matrix Σ_{uu} . In our case,

$$\Sigma_{uu} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 & 0.0 \\ 0.0 & 0.0 & 91.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 3.0 \end{bmatrix} .$$

For each selected sample, the sample regression vector for the errors-in-variables is

$$b(e) = Q_n^{-1}(e) H_n(e)$$

where the rs^{th} element of the (4×4) matrix $Q_n(e)$ is

$$q_{nrs}(e) = \sum_{i=1}^L \sum_{j=1}^2 \sum_{k=1}^{M_{ij}} (x_{ijk_r} x_{ijk_s} - \sigma_{u_r u_s})$$

and the s^{th} element of the $(p \times 1)$ vector $H_n(e)$ is

$$h_{ns}(e) = \sum_{i=1}^L \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} x_{ijk_r} y_{ijk}$$

A consistent estimator for the variance-covariance matrix of $b(e)$ is given by

$$\hat{V}(e) = Q_n^{-1}(e) \hat{G}_n(e) Q_n^{-1}(e),$$

where the rs^{th} element of $\hat{G}_n(e)$ is

$$\hat{g}_{nrs}(e) = 2 \sum_{i=1}^L \sum_{j=1}^2 (\hat{d}_{ij \cdot r} - \bar{\hat{d}}_{i \cdot \cdot r})(\hat{d}_{ij \cdot s} - \bar{\hat{d}}_{i \cdot \cdot s})$$

$$\hat{d}_{ijk_r} = x_{ijk_r} \hat{e}_{ijk}$$

$$\hat{e}_{ijk} = y_{ijk} - \sum_{r=1}^p b_r(e) x_{ijk_r}$$

$$\hat{d}_{ij \cdot r} = \sum_{k=1}^{M_{ij}} \hat{d}_{ijk r}$$

$$\bar{\hat{d}}_{i \cdot \cdot r} = \frac{1}{2} \sum_{j=1}^2 \hat{d}_{ij \cdot r} .$$

The statistic of interest is the "t-statistics"

$$t(b_r(e)) = \frac{b_r(e) - B_r(e)}{s(b_r(e))} , \quad r=1, 2, \dots, 4$$

where $s(b_r(e))$ is the square root of the r^{th} diagonal element of \hat{V}_e .

The data obtained in the 200 samples for each sample design was used for both regression procedures. We present the results of our two experiments in several tables. Table 10 gives for each experiment the mean and variance for the regression coefficients. We observe that the standard errors in design I are approximately $\sqrt{2}$ times the standard errors of the corresponding coefficient in design II. This is to be expected, since the number of primary sampling units in the 12 strata design is twice the number in the 6 strata design. From Table 11, considering the ratio of the estimated bias of 200 sample regression estimates to the estimated standard error of their mean to be distributed as Student's t with 199 d.f., we conclude that the bias is reduced as the sample size increases.

Additional information concerning the frequency distributions of the estimates computed in our Monte-Carlo study is given in Tables 12 and 13 which contains the observed percentiles of the calculated t 's. Examination of Tables 12 and 13 reveals that the distribution for $t(b)$ and $t(b(e))$ agrees more closely with the theoretical t distribution near the median than in the tails. Comparisons of the 1%, 5%, 95% and 99% points for the t statistics in Tables 12 and 13 reveal the effects of increased sample size. For instance, in Table 12 the 5% and 95% points for $t(b_e(4))$ are -2.321 and 2.579 which are considerably higher than the corresponding points for the t distribution with 6 degrees of freedom, ± 1.943 . For these same statistics, the 5% and 95% points in Table 12 are -1.641 and 2.076 as compared to ± 1.782 , the corresponding points for the t distribution with 12 degrees of freedom. These observations suggest that we have underestimated the variances of the sample regression coefficients estimates in small samples, though not by much.

Comparing the results for O.L.S. regression coefficients in Table 14, it is evident that Frankel's calculated t 's for these coefficients are closer to the theoretical t distribution than the ones obtained in our study. One explanation for this is that only urban males between the ages 28-58 were selected for our study. This resulted in decreasing the average number of elements in the sample for designs I and II from 170.3 and 339.5 (as used for Frankel's study) to 61.5 and 124.5 (as used for our study) respectively.

Table 10. Means of 200 regression sample vectors

Number of strata in experiment	Regression coefficients	Least-squares model			
		b_1	b_2	b_3	b_4
6	Population value	8.9289	0.0029	-0.0007	0.0846
6	Means of 200 samples	8.9115	0.0027	-0.0009	0.0812
6	Estimated standard deviation of estimates	0.1136	0.0112	0.0013	0.0308
6	Estimated standard error of mean	0.0080	0.0008	0.0001	0.0022
12	Population value	8.9289	0.0029	-0.0007	0.0846
12	Mean of 200 samples	8.9254	0.0039	-0.0006	0.0842
12	Estimated standard deviation of estimates	0.0724	0.0075	0.0008	0.0228
12	Estimated standard error of mean	0.0051	0.0005	0.0001	0.0016

Errors-in-variables model			
$b_1(e)$	$b_2(e)$	$b_3(e)$	$b_4(e)$
8.9405	0.0053	-0.0006	0.1194
8.9207	0.0055	-0.0008	0.1213
0.1082	0.0116	0.0015	0.0477
0.0076	0.0008	0.0001	0.0034
8.9405	0.0053	-0.0006	0.1194
8.9344	0.0068	-0.0006	0.1225
0.0712	0.0078	0.0009	0.0332
0.0050	0.0005	0.0001	0.0023

Table 11. Estimated bias of regression estimates for 200 replicates

Number of strata in experiment	Least-squares model				Errors-in-variables model			
	b_1	b_2	b_3	b_4	$b_1(e)$	$b_2(e)$	$b_3(e)$	$b_4(e)$
6	-0.0174*	-0.0002	-0.0002*	-0.0034	-0.0198*	0.0002	0.0002	0.0019
12	-0.0035	0.0010	0.0001	-0.0004	-0.0061	-0.0015*	0.0000	-0.0031

*Significant at the 5% level.

Table 12. Comparison of observed percentiles of the calculated t's with the theoretical percentiles for the t distribution with 6 degrees of freedom

Probability in percent	Theoretical percentile for student's t	Observed percentile for t(b)			
		b_1	b_2	b_3	b_4
1	-3.143	-4.841	-3.794	-3.479	-5.315
5	-1.943	-2.616	-2.053	-2.278	-2.650
10	-1.440	-1.855	-1.734	-1.652	-1.773
20	-0.906	-1.070	-1.131	-1.153	-1.314
30	-0.553	-0.695	-0.625	-0.731	-0.823
40	-0.265	-0.392	-0.258	-0.481	-0.434
50	0.0000	-0.057	-0.037	-0.211	-0.222
60	0.265	0.221	0.298	0.153	0.083
70	0.553	0.630	0.632	0.478	0.468
80	0.906	1.236	0.902	0.906	0.982
90	1.440	1.828	1.736	1.874	1.664
95	1.943	2.567	2.898	2.789	2.148
99	3.143	4.418	4.679	5.116	3.638

Observed percentile for $t(b(e))$			
$b_1(e)$	$b_2(e)$	$b_3(e)$	$b_4(e)$
<hr/>			
-4.720	-3.316	-3.329	-5.545
-2.366	-2.055	-2.321	-2.203
-1.667	-1.547	-1.615	-1.811
-0.973	-0.908	-1.110	-1.082
-0.623	-0.515	-0.825	-0.637
-0.328	-0.263	-0.535	-0.395
-0.053	0.016	-0.202	-0.102
0.212	0.262	0.126	0.213
0.428	0.677	0.459	0.551
1.104	0.932	0.826	0.845
1.774	1.564	1.799	1.450
2.849	2.142	2.573	1.666
5.022	3.852	4.890	3.351

Table 13. Comparison of observed percentiles of the calculated t's with the theoretical percentiles for the t distribution with 12 degrees of freedom

Probability in percent	Theoretical percentile for student's t	Observed percentile for t(b)			
		b_1	b_2	b_3	b_4
1	-2.681	-2.442	-3.167	-2.545	-3.278
5	-1.782	-1.777	-1.813	-1.536	-2.306
10	-1.356	-1.364	-1.294	-1.316	-1.440
20	-0.873	-0.975	-0.666	-1.004	-0.961
30	-0.539	-0.554	-0.364	-0.623	-0.451
40	-0.253	-0.195	-0.124	-0.301	-0.145
50	0.000	0.076	0.138	0.032	0.056
60	0.253	0.277	0.492	0.378	0.306
70	0.539	0.640	0.756	0.666	0.694
80	0.873	0.981	1.141	1.043	0.987
90	1.356	1.543	1.709	1.644	1.538
95	1.782	1.976	2.016	1.951	2.028
99	2.681	2.860	2.786	3.300	2.944

Observed percentile for $t(b_e)$			
$b_1(e)$	$b_2(e)$	$b_3(e)$	$b_4(e)$
<hr/>			
-2.694	-2.679	-2.526	-3.222
-1.822	-1.659	-1.641	-1.659
-1.258	-1.273	-1.308	-1.236
-0.842	-0.693	-0.863	-0.796
-0.511	-0.407	-0.542	-0.309
-0.298	-0.090	-0.161	-0.036
-0.024	0.326	0.122	0.195
0.266	0.554	0.412	0.379
0.504	0.784	0.653	0.636
0.938	1.114	0.963	0.930
1.566	1.735	1.516	1.345
1.848	2.112	2.076	1.757
2.824	3.057	3.284	2.428

Table 14. Comparison of observed proportion for calculated $t(b)$ within stated limits to the theoretical proportion for the t distribution

Number of strata in experiment	Intervals	Theoretical proportion	Observed proportion	
			Frankel's study	Our study
6	± 2.576	0.9580	0.9421	0.9350
6	± 1.960	0.9023	0.8733	0.8525
6	± 1.645	0.8489	0.8146	0.8104
6	± 1.282	0.7529	0.7167	0.7037
6	± 1.000	0.6441	0.6029	0.5950
12	± 2.576	0.9757	0.9662	0.9640
12	± 1.960	0.9264	0.9121	0.9103
12	± 1.645	0.8741	0.8496	0.8447
12	± 1.282	0.7760	0.7437	0.7500
12	± 1.000	0.6630	0.6217	0.6100

In summary, the results of this investigation indicate the sample estimates of the multiple regression coefficients have small biases, and the distribution of the t statistics computed for both the O.L.S. and errors-in-variables procedures are well approximated by the theoretical t distribution. In addition, the agreement improves as the number of strata used in the design increase. Our research would have benefitted from even larger numbers of samples. However, calculations required for each sample on a limited budget restricted our study.

VI. SUMMARY

The estimation of regression equations for samples selected from a finite population was investigated. Madow [1948], Erdős-Rényi [1959] and Hájek [1960] have studied the limiting distribution of a properly normalized mean, of a simple random sample selected without replacement from a finite population. Madow's Condition W is a condition on all the moments of the elements of the finite population. Erdős-Rényi's condition may be regarded as a pseudo-Lindeberg condition on the elements of the population. Madow's Condition W was proven to be a stronger condition than Erdős-Rényi's conditions. A more elegant proof is introduced by Hájek using Poisson sampling. By proving the asymptotic equivalence of simple random sampling and Poisson sampling, Hájek obtained the asymptotic normality directly from the Lindeberg Central Limit Theorem. In the three approaches considered, a sequence of finite populations and a corresponding sequence of samples is specified.

Hájek's theorem was extended to the multivariate case as preparation for the study of the asymptotic behavior of regression estimators. A theorem analogous to that of Hájek was obtained for estimated regression coefficients. A sequence of finite populations $\phi_1, \phi_2, \dots, \phi_r, \dots$ with the i^{th} vector element of ϕ_r given by $(Y_{ri1}, Y_{ri2}, \dots, Y_{r,i,p+1})$ is considered. Given that ϕ_r is of size N_r , a simple random nonreplacement sample of size n_r is drawn. The normalized vector

$$Z_{ri.} = (Z_{ri1}, Z_{ri2}, \dots, Z_{r,i,p+1}) ,$$

where

$$Z_{rij} = \frac{Y_{rij} - \bar{Y}_{N_r j}}{\sigma_{Y_{rj}}} , \quad \begin{array}{l} j=1, 2, \dots, p+1; \\ i=1, 2, \dots, N_r \end{array}$$

was introduced. The dependent variable $Z_{r,i,p+1}$ was regressed on $(Z_{ri1}, Z_{ri2}, \dots, Z_{rip})$, $i=1, 2, \dots, N_r$ to give the population regression vector B_r . For the sample, the normalized vector is

$$z_{ri.} = (z_{ri1}, z_{ri2}, \dots, z_{r,i,p+1})$$

where

$$z_{rij} = \frac{Y_{rij} - \bar{Y}_{n_r j}}{s_{Y_{rj}}} .$$

The sample regression vector, b_r , is obtained by regressing $z_{r,i,p+1}$ on $(z_{ri1}, z_{ri2}, \dots, z_{rip})$. Given Hájek-type conditions on the elements $(Y_{ri1}, Y_{ri2}, \dots, Y_{r,i,p+1})$ and existence of $2+2\delta$ moments of Z_{rij} , $j=1, 2, \dots, p+1$, it was proven that the distribution of the properly normalized difference $b_r - B_r$, tended to a multivariate normal.

The limiting distribution of the regression coefficients were also obtained under the assumption that a sequence of finite

populations $\phi_1, \phi_2, \dots, \phi_r, \dots$ of size $N_1, N_2, \dots, N_r, \dots$, where $N_r < N_{r+1}$ was created as a sequence random samples from an infinite multivariate population. It was assumed that the vector elements $(y, x_1, x_2, \dots, x_p)$ of this infinite population had mean $(\mu_y, \mu_{x_1}, \dots, \mu_{x_p})$ and nonsingular covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{xx} \end{pmatrix}.$$

The infinite population regression vector was defined as

$$\beta = \Sigma_{xx}^{-1} \Sigma_{xy}.$$

Letting (y_{ri}, x_{ri}) $i=1, 2, \dots, N_r$ where $x_{ri} = (x_{ri1}, x_{ri2}, \dots, x_{rip})$ be the i^{th} vector in the r^{th} population, we defined the r^{th} finite population vector as

$$B_r = \left(\sum_{i=1}^{N_r} x'_{ri} x_{ri} \right)^{-1} \left(\sum_{i=1}^{N_r} x'_{ri} y_{ri} \right)$$

and the associated r^{th} sample regression vector as

$$b_r = \left(\sum_{i=1}^{n_r} x'_{ri} x_{ri} \right)^{-1} \left(\sum_{i=1}^{n_r} x'_{ri} y_{ri} \right),$$

where n_r denotes the size of the simple random sample selected

without replacement from the finite population. It was shown that the distribution of $n_r^{1/2} (b_r - B_r)$ tended to a multivariate normal distribution with mean zero and covariance matrix $(1-f) \hat{\Sigma}_{xx}^{-1} \theta \hat{\Sigma}_{xx}^{-1}$ where,

$$\theta = E \{ (x.e)' (x.e) \}$$

$$e = y - x \cdot \beta$$

$$x \cdot = (x_1, x_2, \dots, x_p)$$

$$f = \lim_{r \rightarrow \infty} \frac{n_r}{N_r}.$$

A consistent estimator for the variance of $n_r^{1/2} (b_r - B_r)$ was also constructed.

The errors-in-variables cluster model based on the exact mathematical relationships

$$y_{it} = x_{it} \beta_1, \quad i=1, 2, \dots, k; t=1, 2, \dots, n_i$$

where the y_{it} are scalars, the x_{it} are $(1 \times p)$ vectors for the t^{th} element in the i^{th} cluster and β_1 is a $p \times 1$ vector was investigated. The y_t and the elements of x_{it} cannot be observed directly, but only with error. This is a cluster model since clusters are first selected and secondary units are then selected. It was assumed that the error associated with y_{it} could be decomposed into a within cluster component and a between cluster component. The error associated with x_{it} was assumed to be normally and independently distributed as a multivariate normal with mean zero and covariance matrix $\hat{\Sigma}_{uu}$. The X_{it} and Y_{it} were transformed using a Helmert type

transformation (Γ) to F_{it} and G_{it} . This transformation enabled us to define an estimator of β_1 by

$$\tilde{\beta}_1 = (F'F - n^* \mathbb{I}_{uu})^{-1} F'G$$

where $F = \Gamma'X$ and $G = \Gamma'Y$. The estimator, $\tilde{\beta}_1$, is consistent and the limiting distribution of $n^{*\frac{1}{2}} (\tilde{\beta}_1 - \beta_1)$ is normal with mean zero and covariance matrix,

$$\bar{M}_{xx}^{*-1} \sigma_\delta^2 + \bar{M}_{xx}^{*-1} [\sigma_\delta^2 \mathbb{I}_{uu} + (1-c_*) \mathbb{I}_{uv} \mathbb{I}_{vu}] \bar{M}_{xx}^{*-1}.$$

Where

$$\bar{M}_{xx}^* = \lim_{n^* \rightarrow \infty} \frac{f'f}{n^*}$$

$$\sigma_\delta^2 = E (s_{it} - u_{it} \beta_1)^2$$

$$c_* = \lim_{k \rightarrow \infty} \frac{k \sum_{i=1}^k \frac{n_i \sigma_\delta^2 \sigma_w^2}{(\sigma_\delta^2 + n_i \sigma_w^2)^2}}{k}$$

Under the relaxed assumption that the errors associated with x_{it} , y_{it} , possess finite uniformly bounded $4+\delta$ moments, $\delta > 0$, a consistent estimator of the covariance matrix of $n^{*\frac{1}{2}} (\tilde{\beta}_1 - \beta_1)$ was constructed.

Also considered, was an errors-in-variables model wherein the ratio of the response variance to the total variance was known. The model is given by

$$y_t = x_{t.} \beta \quad t=1, 2, \dots, n$$

$$Y_t = y_t + \epsilon_t$$

$$X_{t.} = x_{t.} + u_{t.}$$

where $x_{t.}$ is a p -dimensional vector, Y_t and $X_{t.}$ are observed and the random variables $(\epsilon_t, u_{t.})$ denote errors of measurements. It was assumed that the errors $(\epsilon_t, u_{t.})$ were mutually independent; that $E(\epsilon_t) = 0$ and $E(u_{t.}) = 0$; and that

$$E \begin{pmatrix} \epsilon_t \\ u_{t.} \end{pmatrix} (\epsilon_t, u_{t.}) = \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & D_{uu} \end{pmatrix}.$$

It was also assumed that $x_{t.} \sim \text{NID}(0, \Sigma_{xx})$ independently of ϵ_t and $u_{t.}$. The ratio λ_t ,

$$\lambda_t = \frac{\sigma_{u_t}^2}{\sigma_{x_t}^2} \quad t=1, 2, \dots, p$$

was assumed known. An estimator of β is given by'

$$\tilde{\beta} = (\hat{M}_{XX} - \bigwedge \hat{D}_{XX})^{-1} \hat{M}_{XY}$$

where

$$\hat{M}_{XX} = \frac{1}{n} \sum_{t=1}^n x_t' x_t.$$

$$\hat{M}_{XY} = \frac{1}{n} \sum_{t=1}^n x_t' y_t$$

$$\hat{D}_{XX} = \text{diag } \hat{M}_{XX}$$

$$\Lambda = \text{diag} \left(\frac{\lambda_1}{1+\lambda_1}, \frac{\lambda_2}{1+\lambda_2}, \dots, \frac{\lambda_p}{1+\lambda_p} \right).$$

It was shown that $n^{\frac{1}{2}} (\tilde{\beta} - \beta)$ was asymptotically normal with zero mean and covariance matrix

$$\hat{\Sigma}_{XX}^{-1} V \hat{\Sigma}_{XX}^{-1}$$

where

$$V = (\sigma_v^2 \hat{\Sigma}_{XX} + \hat{\Sigma}_{XV} \hat{\Sigma}_{VX}) - A$$

and A is a $p \times p$ matrix with elements

$$a_{ii} = 2 \Lambda_i^2 \beta_i^2 \sigma_{x_i}^4 \quad i=1, 2, \dots, p;$$

$$a_{ij} = -2\sigma_{x_i x_j} (\Lambda_j^2 \beta_j^2 \sigma_{x_j}^2 + \Lambda_i^2 \beta_i^2 \sigma_{x_i}^2 + \Lambda_i \Lambda_j \beta_i \beta_j \sigma_{x_i x_j}) \quad i \neq j = 1, 2, \dots, p;$$

$$\hat{\beta}_{XV} = -D_{uu} \beta$$

$$\sigma_v^2 = E(v^2) .$$

Given that $(\epsilon_t; u_t.)$ and $(y_t; x_t.)$ had uniformly bounded $4+\delta$ moments, an estimator of the covariance matrix of $\tilde{\beta}$ was constructed.

Finally, a Monte-Carlo study was conducted to study the small sample properties of the ordinary least squares regression estimator and an errors-in-variables regression estimator. The elements of the sample regression vectors were normalized by subtracting the corresponding elements of the population regression vector and dividing the difference by the estimated standard error. The distribution of the resulting statistics, termed "t-statistics", was investigated. It was determined that the distribution agreed with that of the Student's t. The regression vector was approximately unbiased in the Monte Carlo study.

VII. LITERATURE CITED

- Acton, F. S. 1959. Analysis of Straight-Line Data. John Wiley and Sons, New York.
- Adcock, R. J. 1878. "A Problem in Least Squares." Analyst 5:53-54.
- Anderson, T. W. 1958. An Introduction to Multivariate Statistical Analysis. Wiley and Sons, Inc.
- Bailar, B. A. 1969. Recent research in reinterview procedures. J. Amer. Statist. Assoc. 63:41-63.
- Blalock, H. M., C. S. Wells and L. F. Carter. 1969. Statistical Estimation with Random Measurement Error. Pages 75-103 in E. F. Borgatta, ed., Sociological Methodology. Jassey-Bass, San Francisco.
- Bohrnstedt, G. W. and Carter, T. M. 1971. Robustness in regression analysis. Pages 118-146 in E. F. Borgatta, ed., Sociological Methodology. Jassey-Bass, San Francisco.
- Booth, G. D. 1973. The errors-in-variables model when the covariance matrix is not constant. Ph.D. dissertation, Iowa State University.
- Breiman, L. 1968. Probability. Addison-Wesley Publishing Co., Inc., Reading, Mass.
- Brewer, K. R. W. and Ferrier, A. E. 1966. Optimum weighting of outlying observations when sampling from economic populations. Unpublished manuscript. Commonwealth Bureau of Census and Statistics, Canberra, Australia.
- Brewer, K. R. W. and Mellon, R. W. 1973. Problem of analytical surveys or I've already made up my estimator: Don't confuse me with theory. Paper presented at the March 1973 meeting of the Canberra Branch of the Statistical Society of Australia.
- Chernoff, H. 1956. Large sample theory: Parametric case. Ann. Math. Statist. 27:1-22.
- Chung, K. L. 1968. A Course in Probability Theory. Harcourt, Brace and World, New York.
- Cochran, W. G. 1939. The use of the analysis of variance in enumeration by sampling. J. Amer. Statist. Assoc. 34:492-510.

- Cochran, W. G. 1940. Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Statist.* 17:164-177.
- Cochran, W. G. 1963. *Sampling Techniques*. Wiley, New York.
- Cochran, W. G. 1965. The planning of observational studies of human population. *J. Roy. Statist. Soc. A.*, 128:234-260.
- Cochran, W. G. 1968. Errors of measurement in statistics. *Technometrics* 10: 637-666.
- Cochran, W. G. 1970. Some effects of errors of measurement on multiple correlation. *J. Amer. Statist. Assoc.* 65:22-34.
- Cramér, H. 1946. *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- Curtiss, R. F. and Jackson, E. F. 1962. Multiple indicators in survey research. *Amer. J. Soc.*:195-204.
- DeGracie, J. S. and Fuller, W. A. 1972. Estimation of the slope and analysis of covariance when the concomitant variable is measured with error. *J. Amer. Statist. Assoc.* 67:930-937.
- Deming, W. E. 1950. *Some Theory of Sampling*. Wiley, New York.
- Deming, W. E. 1969. Boundaries of statistical inference. Pages 652-670 in N. L. Johnson, H. Smith, *New Developments in Survey Sampling*. Wiley, New York.
- Deming, W. E. and Stephan, F. F. 1941. On the interpretation of censuses as samples. *J. Amer. Statist. Assoc.* 36:45-59.
- Erdős, P. and Rényi, A. 1959. On a central limit theorem for samples from a finite population. *Publ. Math. Inst. Hung. Acad. Sci.* 4:49-61.
- Ericson, W. A. 1969. Subjective Bayesian models in sampling finite populations. *J. Roy. Statist. Soc. B31*:195-234.
- Fisher, R. A. 1928. Book review. *Nature* 156-196.
- Fisher, R. A. 1928. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society* 22:700-725.
- Frankel, M. R. 1971. *Inference from Survey Samples*. University of Michigan, Ann Arbor, Mich.

- Fuller, W. A. 1970. Simple estimators for the mean of skewed populations. Tech. report to the U.S. Bureau of the Census. Iowa State University, Ames.
- Fuller, W. A. 1971. Properties of some estimators for the errors-in-variables model. Paper presented at the Econometric Society Meeting, New Orleans, December, 1971.
- Fuller, W. A. 1972. Notes on the theory of econometrics for Statistics 638, Department of Statistics, Iowa State University, Ames, Iowa.
- Fuller, W. A. 1973. Regression Analysis for Sample Surveys. Paper presented at the first meeting of the International Association of Survey Statisticians. Vienna, Austria.
- Fuller, W. A. and Battese, G. E. 1973. Transformations for estimation of linear models with nested error structure. J. Amer. Statist. Assoc. 68:626-632.
- Fuller, W. A. 1974. Regression Analysis for Sample Surveys. A report prepared for the U.S. Bureau of the Census on work conducted under the Joint Statistical Agreement.
- Gnedenko, B. V. and Kolmogorov, A. N. 1954. Limit Distributions for Sums of Independent Random Variables. Addison-Wesley Publishing Company, Inc.
- Godambe, V. P. and Sprott, D. A. 1971. Foundations of Statistical Inference. Holt Rinehart and Winston. Toronto.
- Graybill, F. A. 1961. An Introduction to Linear Statistical Models. McGraw-Hill Book Company, Inc.
- Hájek, J. 1959. Optimum strategy and other problems in probability sampling. Casspis Post. Mat. 84:387-423.
- Hájek, J. 1960. Limiting distributions in simple random sampling from a finite population. Pub. Math. Inst. Hung. Acad. Sci. 5:361-374.
- Hansen, M. H., W. N. Hurwitz, and W. G. Madow. 1953. Sample Survey Methods and Theory. New York. John Wiley and Sons, Inc. 2 Vols.
- Hartley, H. O. 1959. Analytic studies of survey data. Volume in onora di Corradis Gini. Instituto di Statistics, Rome.

- Isaki, C. T. 1970. Survey designs utilizing prior information. Unpublished Ph.D. Thesis, Iowa State University.
- Johnson, J. 1963. *Econometric Methods*. New York. McGraw-Hill.
- Johnson, N. L. and Smith, H. 1969. *New Developments in Survey Sampling*. Wiley, New York.
- Jöreskog, K. G. 1970. A general method for analysis of covariance structures. *Biometrika* 57:239-251.
- Kalbfleisch, J. D. and Sprott, D. A. 1969. Applications of likelihood and fiducial probabilities to sampling finite populations. Pages 358-387 in N. L. Johnson and H. Smith. *New Developments in Survey Sampling*. Wiley, New York.
- Kempthorne, O. 1971. Probability, statistics and the knowledge business. Pages 671-695, in V. P. Godambe, D. A. Sprott. *Foundations of Statistical Inference*. Holt, Rinehart and Winston, Toronto.
- Kendall, M. G. and Stuart, A. 1961. *The Advanced Theory of Statistics*. Vol. 2. Hafner, New York.
- Kish, L. 1965. *Survey sampling*. John Wiley and Sons, New York.
- Konijn, H. S. 1962. Regression analysis in sample surveys. *J. Amer. Statist. Assoc.* 57:590-606.
- Koopman, T. C. 1937. *Linear regression analysis of economic time series*. Haarlem: DeErven F. Bohn.
- Kummell, C. H. 1879. "Reduction of observed equations which contain more than one observed quantity." *Analyst* 6:97-105.
- Lehmann, E. L. 1959. *Testing Statistical Hypotheses*. John Wiley and Sons, Inc., New York.
- Loève, M. 1963. *Probability Theory*. 3rd ed., Van Nostran Inc., Princeton, N.J.
- Madansky, W. 1959. The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* 54:173-205.
- Madow, W. G. 1948. On the limiting distribution of estimates based on samples from finite universes. *Institute of Mathematical Statistics* 19:535-545.

- Malinvaud, E. 1966. Statistical Methods of Econometrics. North-Holland Publishing Company, Amsterdam.
- Mann, H. B. and Wald, A. 1943. On Stochastic Limit and Order Relationships. Ann. Math. Statist. 14:217-226.
- Moran, P. A. P. 1971. Estimating structural and functional relationships. Journal of Multivariate Analysis 1:232-255.
- Parzen, E. 1960. Modern Probability Theory and its Applications. John Wiley and Sons, New York.
- Mann, H. B. and Wald, A. 1943. On Stochastic Limit and Order Relationships. Ann. Math. Statist. 14:217-226.
- Moran, P. A. P. 1971. Estimating structural and functional relationships. Journal of Multivariate Analysis 1:232-255.
- Press, S. J. 1972. Applied Multivariate Analysis. Holt, Rinehart and Winston Inc.
- Raj, D. 1968. Sampling Theory. McGraw-Hill Book Company.
- Rao, C. R. 1965. Linear Statistical Inference and its Applications. Wiley, New York.
- Rao, J. N. K. 1973. On the foundations of survey sampling. Paper presented at International Symposium on Statistical Design and Linear Models, Colorado State Univ., March, 1973.
- Royall, R. M. 1970. On finite population sampling theory under certain linear regression models. Biometrika 57:377-387.
- Royall, R. M. 1971. Linear regression models in finite population sampling theory. Pages 259-279, in V. P. Godambe and D. A. Sprott. Foundations of Statistical Inference. Holt, Rinehart and Winston, Toronto.
- Scott, A. and Smith, T. M. F. 1969. Estimation in multi-stage surveys. J. Amer. Statist. Assoc. 64:830-840.
- Searle, S. R. 1971. Topics in variance component estimation. Biometrics 27:1-76.
- Sedransk, J. 1965. Analytical surveys with cluster sampling. J. Royal Statist. Soc. B., 27:1121-1139.

- Sukhatme, P. V. 1954. Sampling Theory of Surveys with Applications. Iowa State University Press.
- Sukhatme, P. V. and B. V. Sukhatme. 1970. Sampling Theory of Surveys with Applications. Iowa State University Press.
- Tepping, B. J. 1968. Variance estimation in complex surveys. Proceedings of the Social Statistics Section, Amer. Statist. Assoc. :11-18.
- Tucker, H. G. 1967. A Graduate Course in Probability. Academic Press Inc., New York.
- U.S. Bureau of the Census. 1963. The Current Population Survey. A Report on Methodology. U.S. Government Printing Office, Technical Paper No. 7.
- U.S. Bureau of the Census. 1972. Evaluation and Research Program of the U.S. Census of Population and Housing 1960: Effects of different reinterview techniques on estimates of simple response variance. U.S. Bur. Census, Series ER60, No. 11.
- Villegas, C. 1961. Maximum likelihood estimation of a linear function relationship. Ann. Math. Statist. 32:1048-1062.
- Yates, F. 1949. Sampling Methods for Censuses and Surveys. Griffin, London.
- Walker, H. M. and Lev, J. 1963. Statistical Inference, Holt and Company, New York.
- Wiley, D. E. 1973. The identification problem for structural equation-models with unmeasured variables. In A. S. Goldbager and E. D. Duncan, eds., Structural Equation Models in the Social Sciences. New York, Seminar Press.
- Winakor, G. 1973. Unpublished data from Acquisition and Use of Household Textiles. U.S. Department of Agriculture. Agricultural Research Service Grant 12-14-100-9926.
- Wolter, K. M. 1974. Estimators for a nonlinear relationship. Unpublished Ph.D. dissertation, Iowa State University.

VIII. ACKNOWLEDGMENTS

The author would like to acknowledge the contribution others have made to this work.

My first and most important thanks are to Professor Wayne A. Fuller. Through his patience, encouragement and understanding, this thesis became possible.

I am grateful for numerous discussions about the topic to Professors Chien-Pai Han, Glen Meeden, H. T. David and B. V. Sukhatme.

Also, I am indebted to M. R. Frankel for the Census data tape.

Last but not least, I am truly indebted to my parents for their continued moral support during my stay in Ames.